

12  
b.s.

**LEVEL II**

**RADC-TR-80-366**  
Final Technical Report  
December 1980



AD A 097130

# **DATA COLLECTION ANALYSIS AND TEST**

**Pattern Analysis and Recognition Corporation**

**Dr. Mark R. Nelson**  
**Nori M. Shohara**

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

**DTIC**  
**ELECTE**  
**S** APR 1 1981 **D**  
**B**

DTIC FILE COPY

**ROME AIR DEVELOPMENT CENTER**  
**Air Force Systems Command**  
**Griffiss Air Force Base, New York 13441**

81 4 01 038

This report has been reviewed by the RADC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-80-366 has been reviewed and is approved for publication.

APPROVED:

*Melvin G. Manor, Jr.*

MELVIN G. MANOR, JR.  
Project Engineer

APPROVED:

*Owen R. Lawter*

OWEN R. LAWTER, Colonel, USAF  
Chief, Intelligence & Reconnaissance Division

FOR THE COMMANDER:

*John P. Huss*

JOHN P. HUSS  
Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (IRAA) Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return this copy. Retain or destroy.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

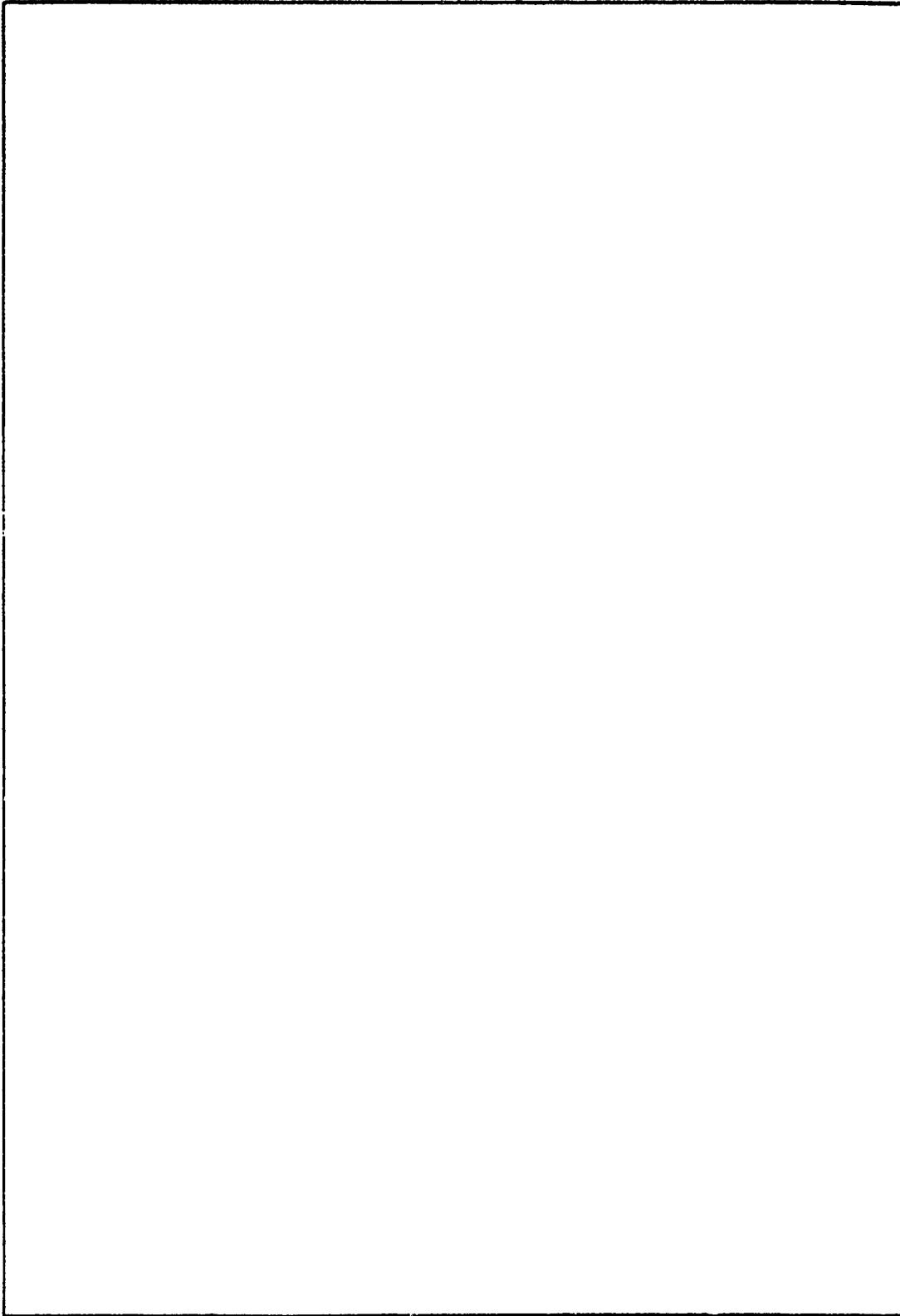
19 REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER RADC TR-80-366	2. GOVT ACCESSION NO. AD-A097160	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) DATA COLLECTION ANALYSIS AND TEST		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report
7. AUTHOR(s) Dr. Mark R./Nelson Mr. Nori M./Shohara		6. PERFORMING ORG. REPORT NUMBER PAR-80-58
9. PERFORMING ORGANIZATION NAME AND ADDRESS Pattern Analysis and Recognition Corporation 228 Liberty Plaza Rome NY 13440		8. CONTRACT OR GRANT NUMBER(s) F30602-79-C-0176
11. CONTROLLING OFFICE NAME AND ADDRESS Rome Air Development Center (IRAA) Griffiss AFB NY 13441		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 63714F 681E0904
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Same		12. REPORT DATE December 1980
		13. NUMBER OF PAGES 134
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) Same		
18. SUPPLEMENTARY NOTES RADC Project Engineer: Melvin G. Manor, Jr. (IRAA)		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Authentication      Signature Verification      Computer Programs Data Base      Software Voice Fingerprint		
ABSTRACT (Continue on reverse side if necessary and identify by block number) This report discusses the results of an effort to determine an experimental procedure for the collection of data bases to be used in testing and evaluating present and future voice, fingerprint, and signature authentication techniques. Areas covered include how much data and what information should be collected, and how it should be collected and stored.		

2/1/81

JCB

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

## TABLE OF CONTENTS

1.0	INTRODUCTION . . . . .	
2.0	QUANTITY OF DATA . . . . .	6
2.1	TYPE I ERROR TESTING . . . . .	6
2.2	TYPE II ERROR TESTING . . . . .	7
2.3	NUMBER OF ENROLLMENT SAMPLES . . . . .	8
2.4	NUMBER OF DATA COLLECTION SESSIONS . . . . .	9
3.0	THE DATA BASES . . . . .	11
3.1	SIGNATURE DATA BASE . . . . .	12
3.1.1	Characteristic Features . . . . .	13
3.1.2	Variations In Features . . . . .	14
3.1.3	Proposed Data CAT System For Signatures . . . . .	16
3.2	FINGERPRINT DATA BASE . . . . .	17
3.2.1	Characteristic Features . . . . .	18
3.2.2	Variations In Features . . . . .	21
3.2.3	Proposed Data CAT System For Fingerprints . . . . .	28
3.3	VOICE DATA BASE . . . . .	30
3.3.1	Characteristic Features . . . . .	32
3.3.2	Variations In Features . . . . .	38
3.3.3	PROPOSED DATA CAT SYSTEM FOR VOICES . . . . .	39
4.0	DATA BASE COMPOSITION . . . . .	49
5.0	CONCLUSIONS AND RECOMMENDATIONS . . . . .	55
	REFERENCES . . . . .	58

### APPENDIX A DATA CAT STATISTICS

A.1	NUMBER OF ATTEMPTS AND SUBJECTS FOR TYPE I ERROR
-----	--

TABLE OF CONTENTS

	TESTING . . . . .	A-1
A.2	NUMBER OF ATTEMPTS AND SUBJECTS FOR TYPE II ERROR TESTING . . . . .	A-17
A.3	NUMBER OF SAMPLES FOR ENROLLMENT . . . . .	A-25
A.4	NUMBER OF SESSIONS FOR EACH SUBJECT . . . . .	A-39

APPENDIX B      PAR SPEECH PROCESSING (PSP) SYSTEM

EVALUATION

This contract was in support of TPO 5B3, Special Projects - Biss. The testing of authentication systems is difficult, expensive, and time-consuming. This effort investigated techniques to develop an automated data base for testing authentication systems. No future work is currently programmed.

*Melvin G. Manor, Jr.*  
MELVIN G. MANOR, JR.  
Project Engineer

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	

## 1.0 INTRODUCTION

As part of the Air Force Base and Installation Security System (BISS) program, Rome Air Development Center (RADC) has sponsored this contract, F30602-79-C-0176, entitled "Data Collection, Analysis and Test" (Data CAT). The purpose of the project was to specify a data base, and its method of collection to be used in testing of present and future voice, fingerprint and signature authentication devices. This report is the final summary of the results of that effort.

Entry control and the associated concept of personal identity authentication have long been of interest to RADC, and are integral parts of the BISS program. A large portion of the effort is devoted to the acquisition of automated entry control systems to provide all levels of security. The diverse requirements of varying applications and levels of security make for a multiplicity of devices and system configurations, all of which require testing and evaluation. The test procedures are expensive and often inconsistent and inadequate. Data CAT is designed to reduce these problems. Specifically, according to the Statement of Work, "The objective of this study is to determine an experimental procedure for the collection of data bases to be used in testing and evaluation of present and future voice, fingerprint, and signature authentication techniques."

A major cost in entry control device testing is the collection of adequate data from test subjects. With Data CAT, this need be done only once, in order to generate the data base. Subsequent testing is performed by reproducing the appropriate attribute from the data. Since the same procedure is followed for every test, results should be consistent and comparable. Furthermore, proper design of the data base will ensure adequate testing.

There were four major issues to be resolved by this effort. First, how much data is required in the data base? For any binary decision making device, there are two types of errors: false rejection and false acceptance. These have been given the names Type I and Type II errors, respectively. We wish to know how much data is required to determine the Type I and Type II error rates to a given confidence. Naturally, we wish to determine the minimum amount of data required, since the cost of collection and storage increases with the quantity of data. This issue speaks to the question of the adequacy of the testing and points out one reason why other procedures were inadequate. Because of a lack of understanding of the statistics of the problem, or to cut costs, inadequate quantities of test data were collected. We have made our determination of data quantity based on a thorough statistical study of the problem.

Specifically, we have determined the minimum total number of test samples and the minimum total number of individuals required to determine a Type I error rate of 1% with 90% and 95% confidence, and a Type II error rate of 2% and .001% with 90% and 95% confidence. We have also made an estimate of the number of samples required for

enrollment, and the number of different sessions required to collect the data.

Second, what information should constitute the data base for each attribute? The answer is, of course, all the information required to reproduce the attribute. This answer is useful though, only in pointing the way to the real resolution of the problem and indeed, to the key risk area of this effort: How to reproduce the attribute. For example, it is not exactly obvious how one should store and reproduce a fingerprint. Optical projection of the image is not adequate because at least one known system requires actual physical contact of the fingerprint ridge on the input sensor [65]. The presence of the ridge changes the index of refraction at the boundary and it is this change which is detected. It has been suggested that the input sensor could be bypassed and its output to the analysis stage could be simulated. This is not acceptable since the sensor is such an important part of the device; its performance must be evaluated also. We propose a procedure which surmounts all these obstacles.

In the case of the voice data base, the difficulty is not in the physical reproduction of the attribute - that can be handled by an amplifier and loudspeaker - the difficulty is in constructing the utterance to be reproduced. The data base must have universal applicability which for voices means that the data must be capable of reproducing an arbitrary utterance. Voice verification devices employ a large variety of utterances for verification and it is not possible to determine a priori which utterances will be required. This fact

dictates that some form of speech synthesis is necessary to reproduce the speech data base. Not only must the utterance be synthesized on some fundamental level, but it also must be recognizably distinct for each subject in the data base. This requirement is indeed a stringent one.

It is clear, then, that the method or procedure used to reproduce the attribute will determine the information to be stored in the data base.

Third, how is the data base to be stored? The resolution of this issue is dictated by the nature of the information to be stored. For instance, analog speech data should be stored on analog magnetic tape. In general, the quantity of data will be fairly large so that some form of archival "off-line" type of storage would seem appropriate. When time comes to test a device, the data could be brought "on-line" to some convenient form. Consider, for example digital speech data. The volume of data is so large that it would not be economical to keep in core memory or even on-line on disk. Digital magnetic tape would be most appropriate. For device testing, the data would be easily transferred from tape to disk, or even read from the tape directly, if random access is not required.

Finally, how should the data be collected? One would like to collect the data in a way that assures its accuracy in representing the population. To do this, one must first determine the population to be sampled, then where to find the subjects, then finally, how to ensure the cooperation of the subjects in obtaining accurate data.

Before pursuing the issues at hand any further, a few general remarks about our approach to the design of the data collection system are in order. Ideally, we would like the collection hardware to be small, portable and inexpensive, as we anticipate collecting data from locales across the nation. Processing and reproduction equipment is not so constrained, so long as the data can be recorded and brought to a central facility. Our system will require a minimum of special purpose hardware, and will be general enough to facilitate expansion and modification.

## 2.0 QUANTITY OF DATA

The first issue addressed was that of determining the amount of data required for Type I and Type II error testing. Recall that a Type I error is a false rejection and that a Type II error is a false acceptance. Derivations of the results presented in this Section appear in Appendix A. Consider first Type I errors.

### 2.1 TYPE I ERROR TESTING

We would like to know the minimum total number of samples required to determine a Type I error rate of  $p = .01$ , or 1% with 90% and 95% confidence. First note that confidences are only defined on intervals about some value. Accordingly, we define an interval of  $\pm .005$  or  $\pm 0.5\%$ , about  $p = 1\%$  which allows a distinction to be made between 1% and 2%. We find then that 1200 test samples will suffice to determine  $p = 1.0 \pm 0.5\%$  with 90% confidence and 1800 test samples gives us 95% confidence in our result. The arguments leading to these results are interesting because they apply to any binary decision with a fixed, constant probability.

To find the minimum total number of test subjects, we first establish that the performance specification  $p = 1.0 \pm 0.5\%$  is the average system performance, not individual average performance. Then assuming the existence of an undisclosed, poorly performing subgroup,

we find that at least 400 subjects must be included in the data base to insure that this subgroup does not unduly affect the results. This notion of subgroups of the population is an important one and will affect the design of the data bases.

Combining the number of samples and individuals tells us that each subject must give at least three to five samples for the test data base for Type I testing.

## 2.2 TYPE II ERROR TESTING

Now let us consider Type II errors. Like Type I errors, we wish to determine the minimum total number of samples and subjects required. The Type II error rates of interest are  $P_1 = 0.02$  or 2% and  $P_2 = 1 \times 10^{-5}$  or .001%. Using the statistics developed for Type I errors, we first define the intervals about  $p_1$  and  $p_2$  to be  $\pm .01$  and  $\pm 0.5 \times 10^{-5}$ , respectively. We find that 800 samples will determine  $p_1 = .02 \pm .01$  with 90% confidence and 1000 samples gives us 95% confidence. The values for  $p_2 = 1 \times 10^{-5} \pm 0.5 \times 10^{-5}$  are  $1.2 \times 10^6$  for 90% confidence and approximately  $1.8 \times 10^6$  for 95%. These are the minimum total number of tests required to determine that the performance meets the specifications.

Using again the notion of undisclosed subgroups, we find that 200 account/intruder pairs are required for  $p = 2\%$  and 399,996 pairs for  $p = .001\%$ . The population of enrolled subjects for Type I testing can be paired for Type II testing. With the restriction that the account and intruder populations cannot overlap, approximately 21 enrolled subjects will form sufficient number of pairs for Type II error of  $2\%$  and 895 for  $.001\%$  error.

### 2.3 NUMBER OF ENROLLMENT SAMPLES

Verification devices require the subjects to first enroll on the system, so enrollment samples must be included in the data base. In keeping with good practice [1], the enrollment samples should be separate from the test samples. How many additional samples should be collected from each subject for enrollment? In general, the answer to this question depends on the dimensionality of the feature space and the complexity of the decision boundary, neither of which are known a priori. An analytic solution is therefore not possible, but it is possible to make a reasonable guess based on current devices. Twenty samples per subject turns out to be a good, conservative figure and indeed, it would seem unlikely that more than twenty samples might be required since an entry control device requiring too large a number of enrollment samples would prove inconvenient to its users.

## 2.4 NUMBER OF DATA COLLECTION SESSIONS

Finally, it is well known that there are certain long-term variations in the attributes under consideration. How many sessions are required to account for these variations? To answer this, assume again that an undisclosed, poorly performing subgroup has emerged as a result of the long-term variations. We can then apply our previous arguments to show that a minimum of 400 collection sessions are required to ensure against the effects of this subgroup. However, if we further assume that the temporal variations are not correlated between subjects, the results for 400 sessions can be inferred from the results for 400 subjects in one session. Therefore, two data collection sessions are required; one to collect enrollment samples, and one to collect test samples. From a study of the long-term variations in the attributes under consideration [22,25,90], it would seem that any period of time longer than four or five days between sessions should be adequate.

In sum, we recommend that data be collected from 400 subjects; their selection and the data collection procedure will be discussed in the sections to follow. The number of samples required for testing is summarized in Table 1. The collection should take place during two sessions.

TABLE 1

Error	Confidence	No. Of Samples	No. Of Samples	Samples Per Subject
Type I, $p=1 \pm .5\%$	90%	1200	400	3
	95%	1800	400	5
Type II, $p=2 \pm 1\%$	90%	800	20	2
	95%	1000	20	3
Type II, $p=.001 \pm .0005\%$	90%	$1.2 \times 10^3$	900	2
	95%	$1.8 \times 10^3$	900	3
Enrollment				20

### 3.0 THE DATA BASES

We would like now to discuss each data base separately and in turn. For each data base, the topics covered will be: The characteristic features of the attribute and their variations; and the proposed system for recording, storing, and reproducing the attribute.

Our aim in studying the variations in the characteristics of the attributes is to be sure that the data base explicitly contains representatives of any known subgroups of the population in proportion with their natural frequency of occurrence. This topic deserves more discussion: The goal of a data base is to represent variability of the known population so that test results will be useful in estimating performance. A data base used to test a device is of limited use if the results do not correspond to the actual performance of the device in the real world, and indeed, this is a problem that plagues any testing program. If accuracy in the test results cannot be guaranteed, certainly precision can be guaranteed by sound design. Such a data base would be useful in comparative evaluation of systems and devices and once experience is gained, correspondence can be made between test results and real world performance.

There is a subgroup of the population which must be included in all the data bases. These are persons with certain physical handicaps. For the voice data base, speech impairments; for the fingerprint and signature data bases, persons with malformed or missing arms, hands, or digits. The reasoning is clear since one would expect that many such persons would have very high Type I error rates, although their Type II error rates would probably be low. In the actual data collection, it would most likely not be necessary to collect data from such persons, and this is reflected in the data base specification.

### 3.1 SIGNATURE DATA BASE

The signature has become the standard means of identity authentication in modern society. It appears on bank drafts and legal documents as proof of the signer's identity. That the signature is subject to forgery is well known and because of this, it serves mainly as a deterrent only to casual imposters. There are really two aspects that a signature provides for identity verification. The first is the static, two-dimensional image itself, signed checks or contracts fall in this category. It does not take a great deal of skill to forge this aspect of a person's signature. The second is the dynamic, ballistic trajectory of the signature as it is produced. Any witnessed signings fall into this category and clearly this is much

more difficult to forge.

It has not been proven conclusively that signatures are unique to an individual. Since signatures are a learned activity, one could certainly imagine that a skilled and dedicated forger could learn to duplicate the exact hand movements of another's signature, right down to the pressure of the dot on an "i", but such effort is hardly practical. Indeed, little is known about the ballistics of signatures or their attributes. Much of what follows is based on our own work and conjecture.

### 3.1.1 Characteristic Features -

We will concentrate our discussion on the ballistic history of the signature rather than the image. The basic information that one might record would be position and pressure (at the tip) as a function of time,  $f(t)$  and  $p(t)$ , respectively. Straightforward differentiation of  $f(t)$  results in the velocity and acceleration of the tip,  $v(t)$  and  $a(t)$ . One may also calculate the curvature,  $\kappa(t)$  or arc-length,  $s(t)$ , or such things as the angle of the pen or the movement of some part of the hand during signing. One may also derive any function in terms of another, for example, velocity and acceleration as a function of position, or arc-length as a function of pressure and so on. This provides a wealth of data from which to extract features.

### 3.1.2 Variations In Features -

All the quantities mentioned in the previous section surely have some natural range. The position,  $f(t)$ , varies over a range of a few centimeters, perhaps up to 10 in the horizontal direction, velocities are on the order of  $10^1$  cm/sec, accelerations are on the order of  $10^2$  cm/sec<sup>2</sup>. Maximum velocities probably occur in the middle of long, slightly curved or straight arcs, and maximum acceleration occurs at points of reversal of direction between two such arcs. (See Figure 1.)

It is difficult to see systematic variations in any of these features that lead to any subgroup of the population. Out of intuition, one would suspect that handedness and possibly gender may systematically affect handwriting. The left-handed mechanics of handwriting are simply different than the right-handed, and this may be evidenced in the production of a signature, if not within the completed image. Everyone has certainly remarked at one time or another that a piece was "written in a woman's hand". These suspicions are borne out by test results of an actual device. [91]

It would be appropriate then, to distribute the handwriting data base according to gender and handedness.

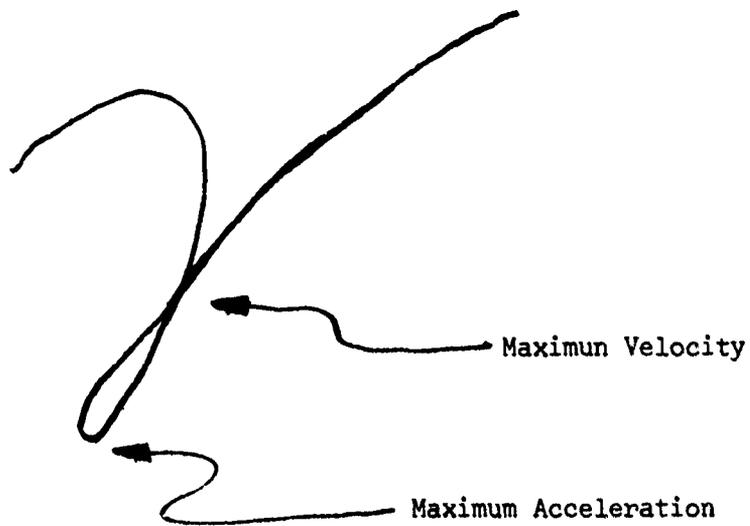


Figure 1  
Typical are in a Signature

### 3.1.3 Proposed Data CAT System For Signatures -

We must reproduce the signature as a ballistic trajectory. The variables required to do this are simply the position as a function of time,  $X(t)$  and  $Y(t)$  and as a substitute for the Z coordinate, the pressure as a function of time,  $P(t)$ . A spatial resolution of .01 in. (.25 mm) should be sufficient. A sampling frequency of 100 Hz results in approximately 1000 data points per signature.

A standard graphics tablet with either a special surface or special pen for pressure sensing would serve excellently for recording the signatures. Data for each subject would be collected and processed in real-time and stored on digital tape. On reproduction, a modified x-y recorder would serve as the output transducer. The recorder would be modified to include a pressure transducer to reproduce the pen pressure. The drawback of this system is that it requires either very special purpose hardware, or a minicomputer for supervising the digitization and recording. This makes for a system that is costly and difficult to transport.

Before any hardware is actually acquired, we recommend a more thorough study of the range of velocities, accelerations and pressures involved in handwriting.

### 3.2 FINGERPRINT DATA BASE

Fingerprints(\*) have had a long history dating back as far as the third century A.D. Evidence from this period suggests that fingerprints were used as seals and identifying marks on some documents. It has only been in the last 100 years, though, that fingerprints were used systematically as a means of identifying people. Their usefulness as an identifying attribute stems from two important qualities. First, fingerprints are unique. No two fingerprints have ever been found to be exactly alike and it is thought by experts that no two ever will be. Cummins [69] gives an estimate of the probability for two fingerprints to be identical as less than one chance in  $10^{43}$ . Since fingerprint patterns are partly controlled by heredity, the assertion that no two are identical is put to the severest test in the case of identical twins. Even in such twins, the prints are at best only similar, not identical. Secondly, fingerprints do not change in form with age unless altered surgically or severely damaged. This has been substantiated by observing the prints of persons taken over intervals of many years [69,71].

(\*) The terms 'print' and 'fingerprint' are used interchangeably and refer to any record of the pattern of lines on the finger, or to the actual pattern on the finger itself. Where a distinction is important, one will be made.

### 3.2.1 Characteristic Features -

Simple examination of the pattern of lines on a finger will reveal all the characteristic features. The pattern consists of ridges (rugae) separated by narrow grooves (sulci), which flow across the finger. The ridges form a global pattern that can be classified as one of three general types: Arches, loops, and whorls (see Figure 2). (The line drawn on the pictures of the loop and whorl are called lines of count; they are not important for this discussion.) The variations are many and it is often difficult to make the distinction between pattern types, but such precision is not necessary for our purposes.

Closer examination (a magnifying glass may prove helpful) reveals more detail. Along the crests of the ridges are tiny impressions that are actually the openings of the sweat pores (the white dots in Figure 2). These are uniquely distributed on every fingerprint and could be used as identifying features by a verification device, but because of their small size, they are difficult to detect and hence are not of practical use. Other local features of the print are obvious. These are the breaks and divergences in the ridge lines that are known as minutiae. These features are of four types: Forks or bifucations, ridge endings, enclosures, and islands (see Figure 3). There are approximately 40 to 200 occurrences of minutia in the average rolled fingerprint.



SIMPLE ARCH

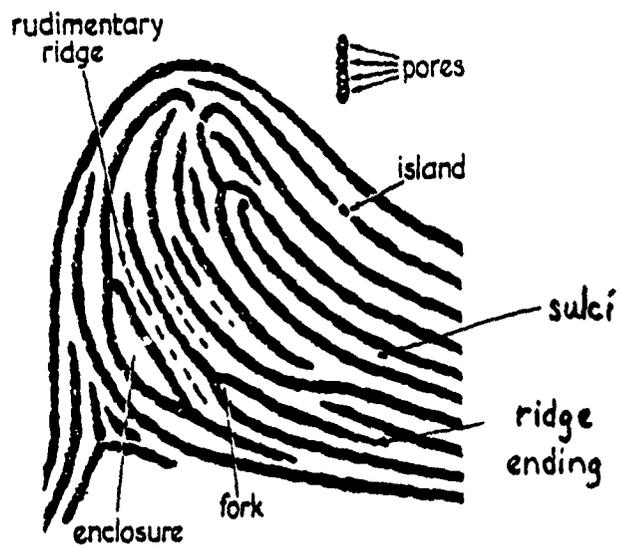


LOOP



WHORL (SYMMETRICAL)

Figure 2.



Details of ridge structure. The rudimentary or secondary ridges have no pores.

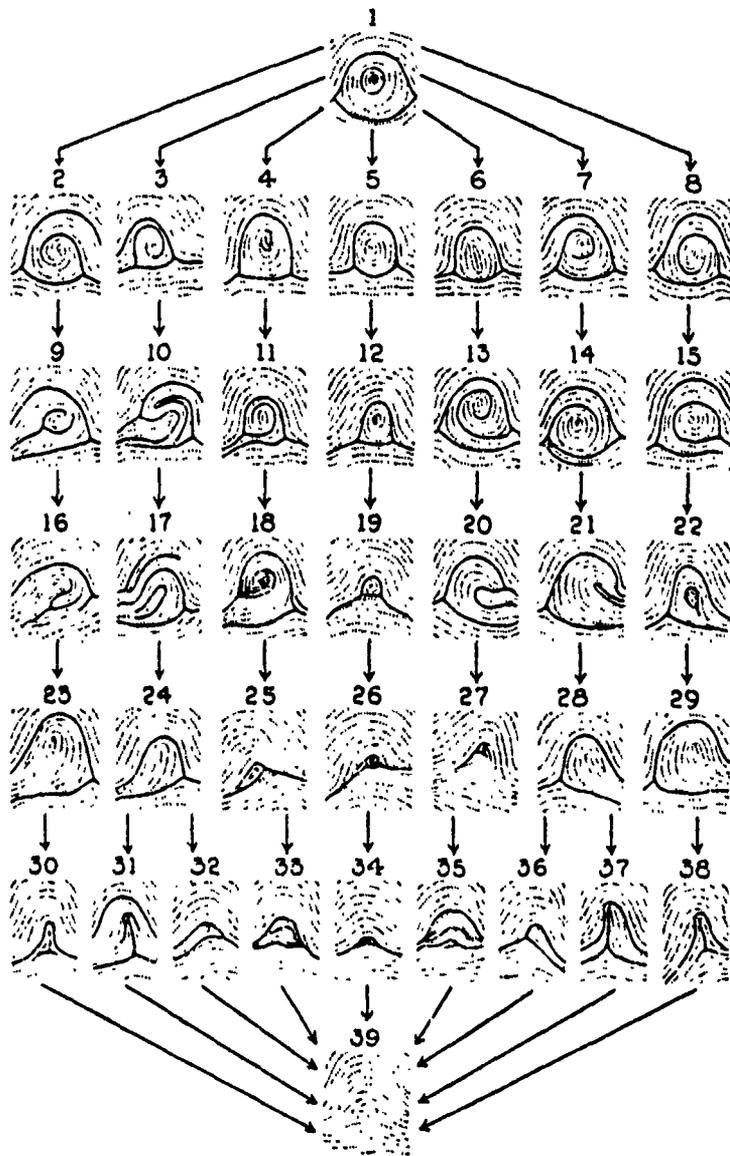
Figure 3.

There are two classes of fingerprint verification/identification devices. The first is based on the local features, the minutia. Basically, the locations of the minutia are extracted and compared with the file prints. The second is generally known as optical correlation and is not quite as successful. Light is passed through transparencies of the test and file prints as they are translated and rotated. The transmittance function is a measure of the correlation between the two (there is an equivalent process in frequency space). [80] We must therefore reproduce both the local features, the minutia, and the global features, the ridge pattern, from our data base.

### 3.2.2 Variations In Features -

Global features vary continuously and a progression of pattern type can be distinguished (see Figure 4). Pattern 1 is an ideal whorl and 39 is an ideal arch. Twenty-four and twenty-eight are loops. Of course the progression can be viewed as going from 1 to 39 or from 39 to 1; no progression in terms of development is implied.

Pattern types are not distributed randomly in the population. The distinction is a statistical one; pattern types occur with varying frequency on each digit of each hand and their occurrence is correlated with race, gender, handedness, and susceptibility to



A "family tree" of fingerprint types.  
 (Modified from Mairs.)

Figure 4.

disease. In general, loops are the most abundant patterns and occur with most frequency on the little finger. Whorls are most common on thumb and ring finger, and the index finger has the highest frequency of arches.

People with certain diseases (e.g., neurofibromatosis, psoriasis, schizophrenia, and so on) tend to have different pattern frequencies than others of similar sex and racial stock [69]. The hypothesis is that some of the same genetic factors that govern fingerprint formation also influence one's susceptibility to disease. Table 2 gives an example of the magnitude of the differences in pattern frequencies for German and Danish schizophrenics.

The difference in pattern type frequency between the control groups of Germans and Danes is typical of what Cummins [69] calls racial variations. He defines his use of the work 'race' thus:

"The sense of 'race' in these examples applies to a group, whether comprehensive or limited, marked by common characteristics traceable to inheritance."

Table 3 is representative of racial variations. We see that in a large sense, Blacks are not distinguishable from Whites, but Orientals appear to have a lower frequency of occurrence of arches.

Also in Table 3, the differences between males and females is shown. In general, females have more occurrences of arches than males. In addition, females are known to have narrower ridges than

Frequencies of Whorls and Arches in Three Independent Series  
of Schizophrenics, Compared with Controls  
From the General Populations

	Germans (Poll)				East Prus- suans* (Duis)		Danes (Møller)			
	Control		Schizo- phrenics		Schizo- phrenics		Control		Schizo- phrenics	
	(845) Male	(776) Female	(232) Male	(545) Female	(416) Male	(356) Female	(86654) Male	(14857) Female	(450) Male	(583) Female
Whorls	33.6%	26.8%	28.5%	28.1%	30.2%	29.6%	29.8%	25.3%	27.0%	26.2%
Arches	4.3	7.6	5.7	6.6	5.2	7.8	5.4	7.5	7.7	8.2

\* The genealogy of all these subjects was traced at least as far as through their grandparents, and East Prussian origin of each generation was established. In the absence of a control, it should be explained that the higher whorl frequencies, as compared with Poll's material, are the expected associate of more frequent whorls in the general population of this territory.

TABLE 2

Pattern-Type Frequencies - Racial Variations

	MALE			FEMALE		
	Arches	Loops	Whorls	Arches	Loops	Whorls
Tobabataks	1.6%	55.4%	43.0%	1.9%	58.5%	39.6%
Koreans	2.3	54.4	43.3	2.8	52.6	44.6
Chinese*	2.5	43.5	54.0	-	-	-
Japanese*	2.7	52.8	44.5	-	-	-
Jews	4.6	53.3	42.1	3.9	52.7	43.4
Danes	5.4	64.8	29.8	7.5	66.3	26.2
Negroes	5.5	65.6	28.9	8.5	63.6	27.9
Germans	6.7	67.1	26.2	8.1	64.9	27.0
Angola Negroes	6.7	67.5	25.8	5.1	64.9	30.0
Dutch	7.7	66.1	26.2	9.6	67.3	23.1
Efe Pygmies	15.9	64.4	19.7	17.0	63.2	19.8

TABLE 3

\* Data for females not available

males; they have  $2.7 \pm .09$  more ridges per centimeter than males (20.7 vs. 23.4).

The fineness of the female ridge structure can have a significant effect on verification device performance [56]. The closeness of the ridges would seem to imply a higher density of minutia on the finger. On this basis then, the gender of the subject is identified as a systematic variable.

Handedness (right or left handed) is related to sex variations in that it tends to cancel them. That is to say that left handed females tend to have the same occurrence of arches as males. For more details concerning variations in fingerprint patterns, see Cummins [69] and Holt [71].

We have yet to specify that pattern type is a systematic variable. Certainly, pattern type frequency does vary with the race, gender, and handedness of the subject, but is the variation significant to the identification problem? We believe not. In the case of minutia based authentication devices, there is no evidence that the occurrence of minutia is correlated with pattern type. In optical correlation, there is no reason to believe that any pattern type is easier to correlate than the others. A second and very practical consideration is that a vast number of subjects would be required if statistically meaningful data is to be collected

representing the various combinations and ranges of pattern type frequency. Note well that we are merely saying that pattern type frequency need not be sampled for explicitly in the data base. Random selection of subjects could result in a data base with pattern type frequencies generally representative of the population.

A variation in fingerprints that is not related to pattern type is physical damage or aberrations. Damage can range from a small cut to complete loss of a digit, hand or arm. Small cuts usually heal and leave no mark visible in the fingerprint. Deeper wounds may leave scars which result in permanent disruption of the pattern. Aside from damage related to disease or accidents, there are certain occupationally related abnormalities. The prints of dishwashers, scrub-women, and workers in lime, plaster and similar substances usually show effects of prolonged exposure to alkali and water. The ridges appear only faintly and are discontinuously printed. These effects disappear once the occupation is abandoned. Such variations should be adequately sampled by random selection from the population.

The maximum size of a rolled fingerprint impression is about 5cm x 5cm. For a pressed print it is about 2.5cm x 5cm. The ridge width varies from .33mm to .75mm; the minutiae are of comparable size. With inked prints, the sulci (light lines between the ridges) are sometimes smeared or partly filled in because of excess ink or pressure, and so vary in size from about .5mm in width to 0mm (i.e., the ridges are indistinguishable). Because of this, very high resolution (.05mm) is needed to read inked fingerprints.

As mentioned previously, the occupation of the subject has some effect on ridge height and there must certainly be some natural variation, but there seems to be no information available concerning this feature of fingerprints.

In sum, beyond the subgroup of the physically handicapped already discussed, we find that the fingerprint data base need include males and females in explicit proportion to their representation in the population. Within those subgroups, random selection of subjects should adequately cover all of the variations mentioned, including occupationally related variations.

### 3.2.3 Proposed Data CAT System For Fingerprints -

The most difficult aspect in designing this system is finding a suitable method of outputting the fingerprint to the verification device. The two methods mentioned earlier, simulating the sensor output and optical projection, have been dismissed as inadequate. We propose to take molds of each of the digits and use these to cast replicas of the digits. The replicas would be stored and 'reproduction' would consist simply of removing them from their storage containers.

The verification device would be tested by manually placing the replica on the input sensor. Data acquisition requires no special transducers, just a spatula for mixing and a mixing pad; there is no data processing, and minimal storage requirements. Accuracy of

reproduction is guaranteed.

After experimentation with materials such as various clays, Silly Putty<sup>®</sup>, and so on, we have found that a suitable material for making the mold is dental impression compound. We recommend KERR<sup>®</sup> PERMLASTIC, regular type III, medium viscosity. The material comes as a catalyst and base, which must be mixed as per instructions. The material is not harmful to skin and is applied directly to the fingertip of the subject, covering it entirely. When dry (approximately 6-8 minutes), the mold is removed and sprayed with a suitable lubricant. Silicone spray lubricant or PAM<sup>®</sup> will suffice. The same compound is then pressed into the mold and allowed to set. When set, the compound has a consistency much like skin, it has a fine sensitivity to detail, and it is non-volatile. The casts are to be made thin so they can be glued to the fingers of a rubber glove on each corresponding fingertip. The gloves should be kept in a cool, dry, dark place to minimize deterioration. To test a device, a technician places his hand in the glove and follows the enrollment and test procedure determined by the device undergoing testing. In this way all individuals are 'reproduced' in the test.

We have produced a small sample of these fingerprints and found the quality to be quite good. The Calspan fingerprint authentication device in the laboratory at RADC was able to register the ridge patterns of the 'reproduced' fingerprint, so we believe this method will prove quite successful. This data base will be simple and inexpensive to collect, maintain, and reproduce and cause minimal user discomfort.

### 3.3 VOICE DATA BASE

The human voice is marvelous in its capabilities and applications. It is the primary mode of human communication. Subtle inflections and rhythms convey the gamut of human emotions and intentions.

To misquote an old adage, how many ways are there to say "I love you"? These same words can be said in all sincerity, mockingly, playfully, derisively, hopelessly, lovingly, and so on, and so on; always the same words, it is the way they are said which conveys the meaning. The extent to which the intended meaning and perceived meaning coincide, however, depends on the skill of the speaker and awareness of the listener. Every Don Juan worth his salt will have command of many modes of expression, and will be able to manipulate the articulators of speech (among other things such as facial expression and hands) to produce the proper cadence, emphasis, and timing to convey a larger message, a more informative message, than just the words might convey. The world about us is full of so many examples of how proper application of the voice means more than saying the right words. A good comedian tells a funny joke; a bad one tells the same joke and it's not funny. The difference is timing, the good comedian would probably say (at least that's what Johnny Carson says: The joke goes something like, "People with good timing either become comedians or parents.").

What is it in speech that allows the same words to say so many different things? There are three ancillary sources for extra information. One is visual cues; hand motions, facial expressions and so on. These play an important role, but we have no interest in them for this project. Another is context. Words uttered in differing contexts change not only their connotations, but even their meaning. Context is of interest here only in how it interacts with the last source; the 'quality' of speech. By 'quality' we mean the emphasis, rhythm, tone, and so on, which a speaker controls in uttering any phrase. These are the factors whose proper manipulation make speech sound natural, and the degree to which this can be done determines the success of one's ability to reproduce speech.

Human beings have the innate ability to manipulate these factors and they employ these abilities with greater or lesser skill. The most talented or influential or persuasive speakers express the ultimate control over not only the quality of their voice, but also the text, visual cues, and context of each phrase. Machines, however, have no such abilities and so must first be given them, then 'taught' to use them. As we have said, this is the key risk area in this effort.

Speech is produced when a pressure, built up in the lungs, is forced past the vocal chords and through the oral and nasal cavities. There are two basic modes of speech. The first is when the vocal chords are held closed. Subglottal air pressure builds until it forces the vocal chords open and a burst of air passes. The vocal chords close once more and the cycle repeats. The period of the cycle

is known as the pitch period. The oral and nasal cavities form a resonant cavity which is excited by the pulse of air coming from the vocal chords, giving rise to what is known as voiced speech. On the other hand, if the vocal chords are held open, the speech is called unvoiced. The excitation of the resonant cavity is furnished by air rushing past a constriction in the vocal tract, giving rise to a noise-like excitation. There is no pitch period for this type of speech.

### 3.3.1 Characteristic features -

We must first decide just what we mean by characteristic features of speech. Do we mean the characteristic features of the speech signal waveform, such as its statistics, frequency structure, or energy content, or do we mean the perceived characteristics of the human voice? There are no compelling arguments that either of those approaches are more appropriate from a technical point of view. Both are equivalent and for the most part, independent. Out of convenience, we choose to consider the perceived speech characteristics. These characteristics are simply those which distinguish dialects of the language in linguistics. This approach is more convenient because of the relatively larger amount of information concerning dialects and also, because of greater ease in screening subjects. If subgroups of the population are identified, say by a particular format structure, then all subjects would have to first be screened by analyzing the format structure of their speech. This does

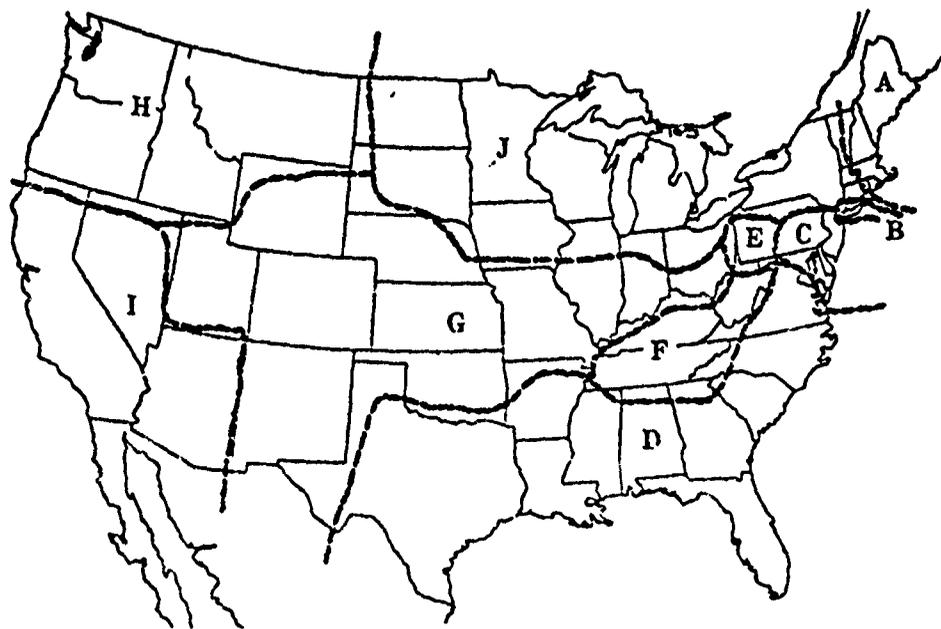
enormously to the effort required to collect the data. There is, however, one feature of the speech waveform which defines easily distinguishable subgroups of the population. This feature is the average pitch frequency which, for females, is about twice as high as males. This fact is known to cause difficulties for verification devices. According to Rosenberg, "The difficulties associated with analysis of female speech are well known. The fundamental problem is the loss of spectral resolution compared with analysis of male speech." [16] The loss of spectral resolution is due to the higher average pitch frequency, leading to more widely spaced harmonics and less information in a given frequency range. We have then the immediate result that the sample group should be divided according to gender.

Dialect is a subjective concept: "Dialects are merely the convenient summaries of observers who bring together certain homogeneities of the speech habits of a group and thus secure for themselves an impression of unity. Other observers might secure different impressions by assembling different habits of the same group." [52] Fortunately, precision in determining an absolute dialect for each subject is not required, we wish only to assure that the sample population represents the major dialectal subgroups. To do this we will attempt to identify the factors which affect one's dialect and from there we can identify the subgroups as those people for which those factors are important.

The first major factor in determining dialect is the mother tongue of the subject: Mother tongue being the first language one acquires. If it is other than American English, then such a person will speak English with a foreign accent. This is not a dialect of American English in a strict sense; it is, however, a variation with which we must contend. Among those with foreign accents we include persons whose mother tongue is British English since British English is spoken differently from American English. We should note here that for our purposes, vocabulary and usage are not important factors in determining dialect. We are concerned mainly with pronunciation, although it is true that such factors undergo similar variations. That is to say, if a person uses a word differently from another, it is more than likely that he pronounces it differently also.

The next most important element in determining dialect is the region of origin of the speaker. These influences result from local, regional variations in speech and are established in a child by adolescence. It is not possible to draw definitive regional boundaries, and every expert will propose slightly different ones, but as we have said, precision is not required. The map in Figure 5 gives an acceptable subdivision of the United States into ten linguistic regions.

In general, socioeconomic status has a profound affect on the nature and extent of linguistic variation. A typical example is given in [47] for the occurrence of postvocalic 'r' absence:



Map showing the major regional speech areas:  
A: Eastern New England; B: New York City; C: Middle Atlantic; D: Southern; E: Western Pennsylvania; F: Southern Mountain; G: Central Midland; H: Northwest; I: Southwest; J: North Central.

Figure 5.

Socioeconomic Class	Mean % 'r' Absence
upper middle	20.8
lower middle	38.8
upper working	61.3
lower working	71.7

The middle classes show more homogeneous speech habits across regional boundaries, the lower classes exhibit the regional peculiarities more strongly, though this may be less true in the South and Southern Mountain regions where upper and middle class speakers speak a fairly strong regional dialect. According to Wolfram and Fasold [47], the best indicators of socioeconomic status are education, occupation, income (both source and amount), house type, and dwelling area.

There is a dialect known as Vernacular Black English which is common only among lower class urban blacks. This fact brings us to the question of the effect of the speaker's race or ethnic background on his speech. It has been proposed that there are physical features of vocal tracts that differ according to race; this especially in connection with Vernacular Black English. However, this proposal is not generally accepted by linguists and comparative anatomical studies do not support it. Aspects of linguistic behavior that are highly correlated with race (more specifically, highly correlated with being black) are due to factors which cause the black community to be highly segregated socially from general American influence. No other racial

or ethnic group, except of course those groups whose mother tongue is not English, show any systematic variation. Studies have shown that the dialect of Puerto Ricans in New York City is affected most by their peer group contacts, even when there is strong parental influence in other directions [47]. The persistence of Vernacular Black English is easy to understand in this light; people growing up in the urban black community are affected most by their peers and since urban black neighborhoods are inevitably segregated, those peers speak Vernacular Black. The dialect is perpetuated by the same social forces that perpetuate segregation. The influence of peer groups is far reaching. Quoting Wolfram and Fasold [47], "Although interference from a foreign language may be quite obvious in the speech of first-generation immigrants, straightforward interference from another language is of little or no significance for the second and third-generation immigrant." This is because English language skills are acquired through peer group contacts. This is indeed an important point. One may at first suspect that not only persons whose mother tongue is not English should be accounted for, but also those who grew up in households where the predominant language was not English should be accounted for. Fortunately, we see that this is not the case since the mechanism of peer group influence tends to homogenize speech patterns within a given community. For our purposes, speakers of Vernacular Black English form a recognizable subgroup of the population.

### 3.3.2 Variations In Features -

What sorts of variations are there among the different dialects? Besides variations in vocabulary and usage, the major difference is in pronunciation, principally of the vowels or, more generally, voiced sounds. Referring to the map in Figure 5, the variations seen in regions G, H, I, and J are subtle. In fact, many linguists classify inhabitants of these regions as all speaking one dialect known as general American English. Speakers from the southern region tend to slur and elongate vowel sounds. This changes the rhythm of the speech and gives rise to the Southern drawl. Residents of the New England area tend to nasalize vowels which results in the "New England twang". Persons from central Pennsylvania have a unique dialect known as Pennsylvania-Dutch. It results from German (Deutsch) influence rather than Dutch influence, as it first might be thought, and is marked by confusion of sounds such as 'b' and 'p', 'd' and 't', and others. There is, of course, much richer regional variation than outlined here, however, the details are not important.

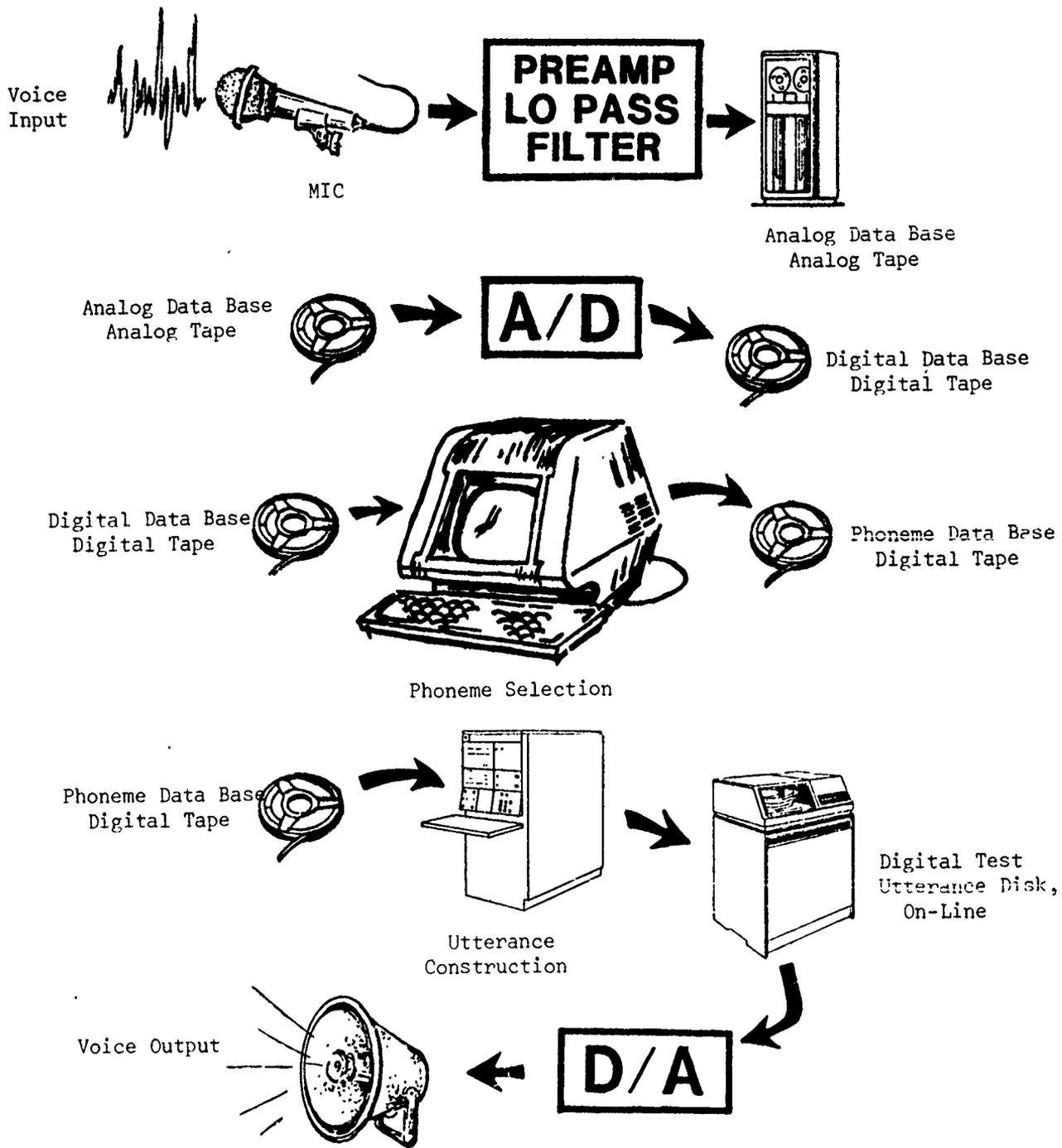
There are variations in the speech signal which are important. The maximum frequency range of the human voice is approximately 50-6000 Hz, although there is very little information in the higher frequencies. The dynamic range of the human voice is 30 - 40 dB [55].

### 3.3.3 PROPOSED DATACAT SYSTEM FOR VOICES -

As we have said above, it is clear that some form of speech synthesis is required in reproducing the speech data base. The method chosen for the synthesis will determine the details of the system, but the general form it will take is clear. A set of phrases will be specified which contain all the phonetic events required for synthesis. These phrases will be recorded for each subject on analog tape. Thus, the data collection equipment is inexpensive and portable. The analog tape is then brought to the computer facility where it is digitized. Phonemes are then selected to form the phoneme data base. The term phoneme is used here, not in the linguistic sense, but in the broad sense meaning the fundamental building blocks the speech will be constructed from. From the phoneme data base, the test utterances required by the verification device are constructed, then converted to analog form and used for the test. This procedure is diagrammed in Figure 6.

The analog data base is stored on analog tape, the digitized speech is stored on digital tape, as is the phoneme data base. The digitized test utterance can be held on-line on disk or off-line on digital tape.

What exactly is required from our speech synthesis? We must reproduce a speaker: We must collect data from a subject and use it to reconstruct his speech. One might call it speaker synthesis. It



## SPEECH PROCESSING SYSTEM

FIGURE 6

was clear from the outset that this was not a trivial task. From a linguistic point of view, speech is not characterized well enough on an individual level such that all of a person's speech habits, his personal dialect, can be known from some limited set of data. From a technical point of view, state-of-the-art capabilities allowed the synthesis of natural sounding speech. To generate speech that was not only natural, but sounded like some individual, would take another advance in the state-of-the-art. There are some aspects of this problem, however, which allow for compromise. The verification device under test does not have to verify that the reproduced voice be the same as the original speaker. It is required only to distinguish utterances constructed from one phoneme set from those constructed from all other sets, with the specified accuracy. This eases the requirements somewhat. We do not have to reproduce a set of human speakers, we have only to produce a set of voices whose characteristics are representative of the population. The specifications then, for the quantity and type of data to be collected are crucial since it is here that the data base makes contact with the real world. Additionally, the psychology of entry control argues that the users will grow accustomed to the system and will learn, subconsciously, to repeat the verification phrase in such a way as to gain access. Such a system is a classic example of what psychologists call operant conditioning, with the reward being successful access. Untold numbers of rats have learned to run mazes in just this fashion. Experience with existing systems supports this supposition. In fact,

an entry control device may actually use different decision strategies for users new to the system and those experienced with it [36]. The new users are judged less strictly while they adjust to and learn the system. What this all means is that the tremendous variety and richness of which speech is capable will not be present in a verification situation. Just as one would expect, a typical user of the system will not expound grandly his verification phrase one day, then coo softly on the next. He would, in general, recite it in the same pat manner as he did originally during enrollment. Since the context of the situation and phrase never change, no variation in pronunciation should be expected due to context, and finally, of course, visual cues or motions are of no consequence. In short, we now find that it is not necessary to reproduce a specific speaker, nor is it necessary to reproduce all aspects of speech and vocal expression in order for this data base to meet its goals. It will suffice for us to simply produce from each data set from each speaker, the utterance required in a natural sounding voice.

Let us now discuss our actual speech synthesis system. It is based on the source filter model of speech production as depicted in Figure 7.

In this model it is assumed that the exciting source, the vocal chords or a vocal tract constriction is linearly separable from the remainder of the vocal tract, which acts like a filter. As we have said before, the exciting source is either a pulse train or white noise. The filter can be any appropriate filter either real or modeled. Of course, because of the flexibility available, this system

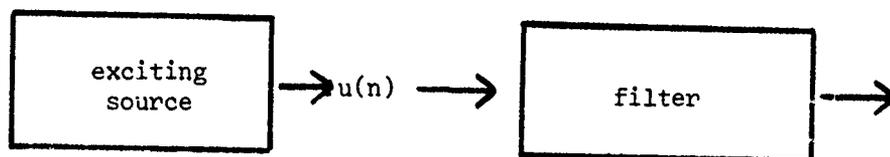


Figure 7.

is ideally simulated on a general purpose digital computer. Well established techniques exist for estimating the filter characteristics from the speech signal. Assuming the filter can be adequately modeled by an all pole filter of moderate size, linear predictive coding technique turns out to be very useful [32].

Let the discrete time series output of the system be  $s_n$ , the previous outputs  $\tilde{s}_n$  and inputs  $u_n$ , then the system can be modeled by:

$$s_n = \sum_{k=1}^p a_k \tilde{s}_{n-k} + G \sum_{\ell=0}^q b_\ell u_{n-k}, \quad b_0=1$$

$G$  is the gain factor. Taking the  $z$  transform, the transform function of the filter is given as:

$$H(z) = \frac{S(z)}{U(z)} = G \frac{1 + \sum_{\ell=1}^q b_\ell z^{-\ell}}{1 + \sum_{k=1}^p a_k z^{-k}}$$

where  $S(z)$  and  $U(z)$  are the  $z$  transforms of the output and input. This is known as the pole-zero model. The transfer function can be estimated to any desired degree of accuracy if all  $b_1 = 0$ , then:

$$H(z) = G \frac{1}{1 + \sum_{k=1}^p a_k z^{-k}}$$

The problem is to estimate the  $a_k$ , the linear predictive coefficients, and to choose  $p$  such that the filter is determined to the desired

degree of accuracy. Procedures for doing this are well established in the literature, and one of the most common is known as the method of least squares.

Assume the input  $u_n$  to the system is not known. The output  $s_n$  can then only be approximated by:

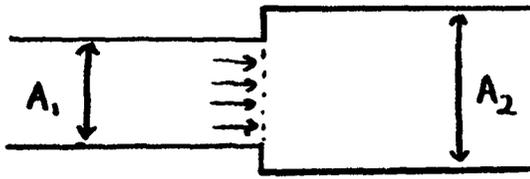
$$\hat{s}_n = \sum_{k=1}^p a_k s_{n-k}$$

The error  $e_n$  between the actual value and the predicted value is simply the difference:

$$e_n = s_n - \hat{s}_n = s_n - \sum_{k=1}^p a_k s_{n-k}$$

$e_n$  is also known as the residual, and the  $a_k$  can be determined by minimizing the mean total squared error. The result is a set of  $p$  simultaneous equations in  $p$  unknowns and is the same for a deterministic or random signal. Computationally economical methods are known for solving these equations and from them we have chosen to implement the auto-correlation method.

An added benefit from the auto-correlation method is a secondary set of coefficients known as the partial correlation or reflection coefficients. The term reflection coefficients arises from transmission line theory where the reflection coefficients are actually those of the boundary between two regions of differing impedance with a plane wave normally incident at that boundary. In the case of speech, the natural transmission line is an acoustic tube made up of equal length sections of constant but differing cross-sectional area.



The reflection coefficients,  $k_n$  are related to the cross-sectional area  $A_n$  by:

$$k_n = \frac{A_n - A_{n+1}}{A_n + A_{n+1}}$$

The analog speech is digitized at 12.0 kHz, giving an effective bandwidth of 6.4 kHz. We use a 20 ms processing frame length which corresponds to 256 data points per frame. Each frame of digitized speech is encoded using linear predictive coding. The data is pre-emphasized then windowed with a 256 point Hamming window, then the voicing, pitch period, LPC coefficients, reflection coefficients and cross-sectional areas are extracted for each processing frame. Each phoneme is represented by one frame of data. Phoneme selection is interactive and aided by waveform displays and automatic phoneme recognition. The operator must make the final determination of which frame represents the desired phoneme. A library containing all the required phonemes will be assembled for each subject.

To construct a new utterance, the operator specifies a string of phonemes along with a relative gain and pitch and a duration. Because pitch and duration are under operator control, he is responsible for obtaining the proper prosody. The difficulty with this approach to speech synthesis lies in handling the transition from one phoneme to

the next. The implicit assumption is made that connected speech can be modeled as a series of steady state phonemes, reasonably invariant from occurrence to occurrence, which are connected by smooth transitions. The speech synthesis program must calculate the transitions.

Others have tried this approach and have met with little success because they calculated transitions by interpolating between successive sets of LPC coefficients. There is no reason to believe that this scheme has any physical basis and indeed, If one looks at the time history of the LPC coefficients, one finds they do not change smoothly. The solution is to interpolate on a set of physically meaningful coefficients, the cross-sectional areas. An extensive survey of the cross-sectional areas in natural speech has resulted in interpolation rules. Our experiments in this area show that this method does work. We have constructed a set of phrases taken from the Texas Instruments Automatic Speaker Verification System [38]. These phrases have good, natural sounding quality and can be recognized as the voice of the original speaker.

The operator has the complete capability to audition the constructed utterance and make changes he deems appropriate. The operator should have expertise in dialectology so that he will be useful in segmentation and construction.

Each subject in this data base has three permanent data sets. The analog recording of the original passage on audio tape, the digitized version of this, and the phoneme library, which is also digital. These are most economically stored on magnetic tape, the format depending on the particular computer installation on which the processing was done. In our research, a DEC PDP 11/70 was used. The digitized data is stored in 512 byte (256 data point) blocks, as unformatted 2 byte integers. The phoneme data base is stored as unformatted 4 byte real data. For device testing, the phoneme library for each subject is brought on-line from tape, the test utterances are constructed, then stored on digital tape in the same format as the original digitized utterances. After the processing is completed, these can be played out to the device through the D/A interface, amplifier and loudspeaker.

#### 4.C DATA BASE COMPOSITION

We have discussed so far the quantity of data required, its form and methods of storage and reproduction. We will now describe the actual data collection.

The most economical way to collect the data base will be to use portable equipment which can be brought to the collection site. We recommend the data be collected at U.S. military installations since all subjects required are likely to be found there. Each subject will be sampled for his fingerprints, signature and voice. Care should be taken in screening subjects and to insure accurate data. Both civilian and military personnel, officers, and enlisted men should be included in the population.

We have identified subgroups of the population for each attribute and the sample should be assembled accordingly. The sample should be half male and half female. Each of these groups should then be divided according to mother tongue, then region of origin. Within the smallest subdivisions, subjects should be drawn at random (see Figure 8). This satisfies the requirement for the voice and fingerprint data base, but the signature data base requires it be distributed according to handedness. Further subdivision of the population is undesirable since small numbers of persons in a subdivision would not lead to statistically meaningful results. Rather than increase the sample

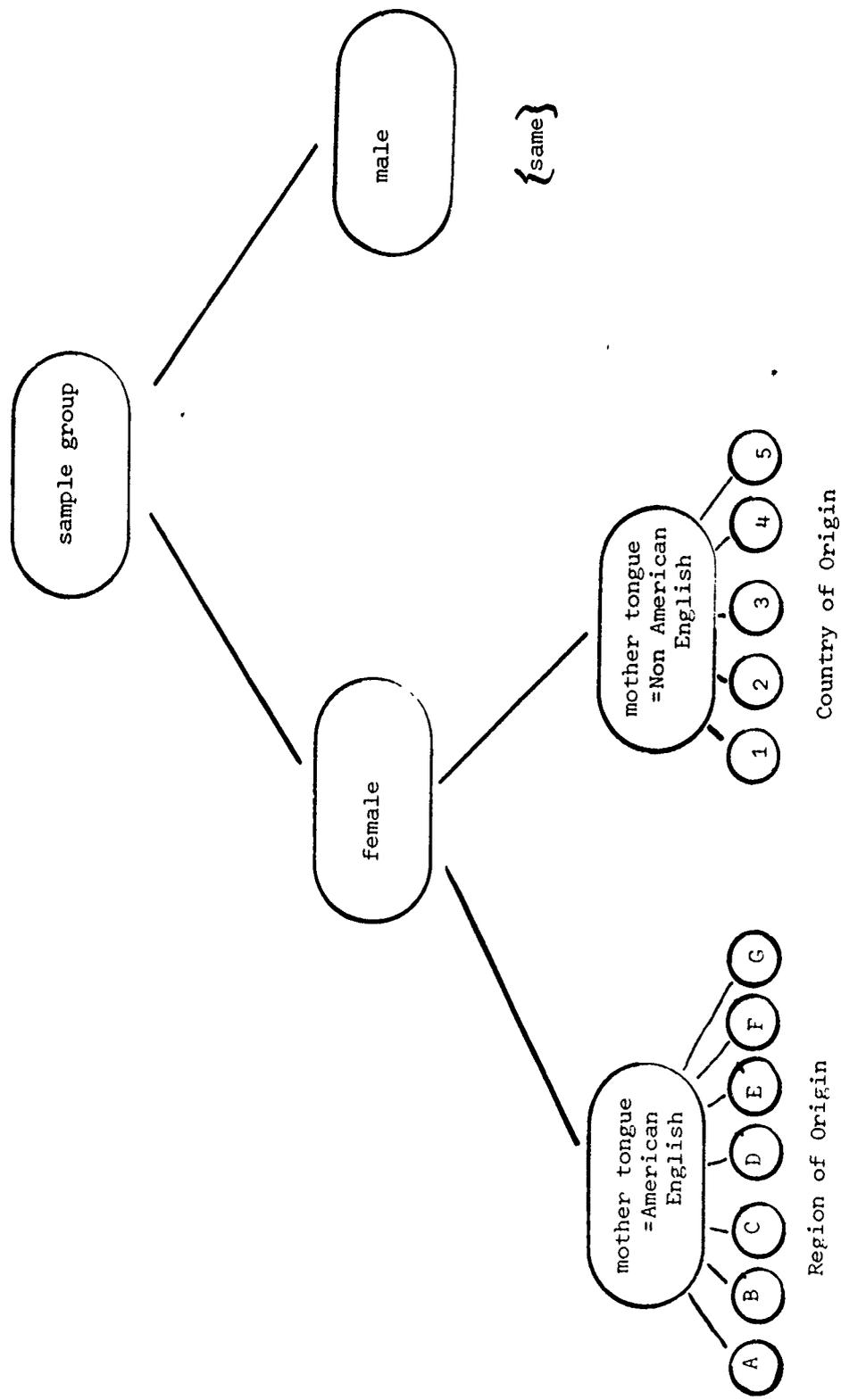


Figure 8.

size, we believe that the random draw will adequately sample the right and left handed populations. Table 3 gives a breakdown of the U.S. population into the subgroups we have identified. Of the 400 subjects, 200 will be male and 200 female. The last column in Table 4 gives the number of male and female subjects required for each region.

The entire population should be characterized in terms of the important variables that we have defined: gender, mother tongue, and region of origin. Only those raised from birth through adolescence in one region shall be considered as true members of that subgroup. Others may speak with a dialect reflecting the influence of two or more different regions; similarly with mother tongue. Once so divided, names can be drawn at random and the named person can be asked to participate in the study. The voluntary participation of the subject should give some confidence that he will be cooperative. So as not to stretch our confidence in human nature too far, we suggest a small monetary compensation may buy a little more cooperation.

Once the subject has been secured, a short briefing explaining the purpose of the project should be given to orient the subject and to give him time to relax. Every subject should be reassured that the information collected in this study will be used only for the stated purpose and will not be circulated without his permission. The fingerprints, being unaffected by the emotional state of the subject, should be collected first. The signatures (signing being a very natural act) should be collected next, then finally the voice data

Region	American English	%	Out of 160 Samples
Eastern New England	5,025,180	3.7	6
New York City	6,631,491	4.2	7
Mid Atlantic	12,676,239	8.0	13
Southern	37,007,323	23.5	37
Western Pennsylvania	4,572,180	2.9	5
Southern Mountains	8,445,017	5.3	9
Mid Central	25,056,602	15.8	25
Northwest	5,862,160	3.7	6
Southwest	15,575,619	9.9	16
North Central	30,257,947	22.9	36
<b>Total</b>	<b>158,049,799</b>	<b>100.0</b>	<b>160</b>

Mother Tongue	Non American English	%	Out of 40 Samples
Spanish	7,823,583	17.9	8
German	6,023,054	14.0	7
Italian	4,144,315	9.5	5
French	2,598,400	6.0	3
Polish	2,437,900	5.6	3
English	1,697,025	3.9	3
Yiddish	1,593,993	3.7	2
Russian	334,565	0.8	1
Other	8,149,266	18.7	8
Not Reported	8,764,358	20.1	- (*)
<b>Total</b>	<b>43,637,305</b>	<b>100.0</b>	<b>40</b>

TABLE 4

(\*) Subjects credited to 'unreported' were distributed evenly among all other categories (one additional subject for each).

base. The voice recording should be done in a sound booth or a quiet room. The subject should be given time to familiarize himself with the text to be recited, and any ambiguities or questions should be cleared up prior to recording. The recording should not be rushed and the subject should be allowed to pause if desired. All precautions should be taken to insure recording the subject in as natural a state as possible. Figure 9 is a list of equipment required for recording the voice data base.

List of Data CAT Speech Processing Hardware

Qty	Item	Specification/Recommendations
1	Microphone/Preamp	Frequency response 50-6000 Hz cardoid condenser/FET preamp. windscreen, associated hardware AKG CK1 Condenser Mic C451E FET Preamp W3 Windscreen
1	Linear Audio Amplifier	Frequency response 50-6000 Hz Variable Gain S/N > 60db
1	Bandpass Audio Filter	Low cut 50 Hz High cut 6000 Hz S/N > 60dB
1	Analog Audio Tape Recorder/Player (2 or 4 track)	Frequency Response 50-6000 Hz S/N > 60dB THD < 0.5%
1	Audio Loudspeaker	Frequency Response 50-6000 Hz High Efficiency, 4 or 8 Ohms
1	Computer w/Analog Interface	Digital Tape - Large Disk > 12 bit A/D, D/A > 12000 Hz Sampling Rate DEC PDP 11/70 w/ LPA11-K RPC4, TU16
1	Graphics Terminal	Waveform Display Tektronix 4014
	Miscellaneous	Cables; Connectors; Magnetic Tape

Figure 9

## 5.0 CONCLUSIONS AND RECOMMENDATIONS

We believe that the Data CAT approach is basically sound. The relatively high expense of collecting the signature data base, weighed against its possible uses, leads us to believe that this effort would not be cost effective. The fingerprint data base is extremely easy and inexpensive to collect and could prove very useful in testing not only fingerprint identification devices, but also fingerprint recognition and classification devices. Since this data base need only be subdivided by gender, the collection could take place in any population center and large number of fingerprints could be included in the data base at very low cost.

The voice data base has a moderate initial cost due to the acquisition of required equipment, and the screening of subjects is more costly, but the potential benefits are very great considering the growing field of speech identification and recognition. As with the fingerprint data base, the speaker synthesis system and voice data base can be useful testing both speaker identification and speech recognition devices.

Since this is a new application of new technology, it may be wise to proceed cautiously in its development. The cost of the actual data collection will obviously far outweigh the cost of equipment and software required for the processing. However, the capability to

acquire the data base would not be that costly. The speaker synthesis unit, consisting of a host computer, analog interface, graphics capability, software, and associated analog equipment could be procured and a small data base collected for minimal cost. This would allow the user the opportunity to prove the technology under laboratory conditions and also establish baseline performance and real world performance, thus avoiding the typical pitfall of precise but inaccurate testing.

As an added bonus, software and techniques developed under contract F30602-79-C-0226, known as UNITRANS, also sponsored by RADC and recently completed by PAR [92] could be easily integrated into the speaker synthesis unit, providing the user with a virtually unlimited number of synthetic speakers and virtually unlimited speech synthesis capability.

As with any good laboratory tool, the uses of such a system are innumerable. It could be used in testing speaker verification devices, as per its original intent, speech recognition devices, voice communication and bandwidth compression systems, computer simulations of the above, and so on.

## REFERENCES

- [1] D.H. Foley, "Considerations of Sample and Feature Size", IEEE Transactions on Information Theory, Vol. IT-18, No. 5, September 1972.
- [2] P. Cumiskey, N.S. Jayant, J.L. Flanagan, "Adaptive Quantization in Differential PCM Coding of Speech", Bell Syst Tech J., Vol. 52, pp.1105-18, September 1973.
- [3] N. S. Jaynant, "Adaptive Delta Modulation With a 1-Bit Memory", Bell Syst Tech J., Vol. 49, pp.321-42, March 1970.
- [4] B.S. Atal, M.R. Schroeder, "Adaptive Predictive Coding Speech Signals", Bell Syst Tech J., Vol. 49, pp.1973-83, October 1970.
- [5] J.L. Flanagan, "Computers That Talk and Listen", Proc IEEE, Vol. 64, pp.405-15, April 1976.
- [6] L.R. Rabiner, R.W. Schafer, "Digital Techniques For Computer Voice Response: Implementations and Applications", Proc IEEE, Vol. 64, pp.416-33, April 1976.
- [7] L.R. Rabiner, R.W. Schafer, J.L. Flanagan, "Computer Synthesis of Speech by Conctenation of Formant Coded Words", Bell Syst Tech J., Vol. 50, pp.1541-58, 1971.
- [8] B.S. Atal, S.L. Hanauer, "Speech Analysis and Synthesis By Linear Prediction of the Speech Wave", J Acoust Soc of Amer, Vol. 50, pp.637-55, 1971.
- [9] J.P. Olive, "Fundamental Frequency Rules for the Synthesis of Simple Declarative English Sentences", J Acoust Soc of Amer, Vol. 57, pp.476-82, 1975.
- [10] S.K. Das, W.S. Mohn, "Pattern Recognition in Speaker Verification", Proc FJCC, Vol. 35 (Las Vegas, Nev.), 1969.
- [11] R.C. Lummis, "Speaker Verification: A Step Toward the 'Checkless' Society", Bell Lab Rec, Vol. 50, pp.254-59, September 1972.

- [12] A.E. Rosenberg, "Listener Performance in Speaker Verification Tasks", IEEE Trans Audio Electro., Vol. AU-21, pp.221-25, 1973.
- [13] G. Doddington, "Speaker Verification", RADC, Griffiss AFB, NY, Tech Report RADC-TR-U1-963700-F, April 1974.
- [14] A.E. Rosenberg, M.R. Sambur, "New Techniques For Automatic Speaker Verification", IEEE Trans Acoust Speech Sig Proc, Vol. ASSP-23, pp.169-76, 1975.
- [15] R.C. Lummis, "Speaker Verification by Computer Using Speech Intensity for Temporal Registration", IEEE Trans Audio Electroacoust, Vol. AU-21, pp.80-89, 1973.
- [16] A.E. Rosenberg, "Automatic Speaker Verification: A Review", Proceedings IEEE, Vol 64, pp.460-74, April 1976.
- [17] N.S. Jayant, "Adaptive Quantization With a One-Word Memory", Bell Syst Tech J., Vol. 52, pp.1119-44, September 1973.
- [18] J.L. Flanagan, C.H. Coker, L.R. Rabiner, R.W. Schafer, N. Umeda, "Synthetic Voices for Computers", IEEE Spectrum, Vol. 7, pp.22-45, October 1970.
- [19] S.K. Das, W.S. Mohn, "A Scheme for Speech Processing in Automatic Speaker Verification", IEEE Trans Audio Electroacoust, Vol. AU-19, pp.32-43, March 1971.
- [20] E. Bunge, "Automatic Speaker Recognition by Computers", 1975 Carnahan Conference.
- [21] J.E. Paul, Jr., Ph.D., et al., "Development of Analytical Methods for Semi-Automatic Speaker Identification System", 1975 Carnahan Conference, pp.52-64.
- [22] W. Endress, W. Bambach, G. Flosser, "Voice Spectrograms as a Function of Age, Voice Disguise and Voice Imitation", J Acoust Soc of Amer, Vol. 49, pp.1842-49, 1971.
- [23] H. Hollien, W. Majewski, P. Hollien, "Perceptual Identification of Voices Under Normal Stress and Disguised Speaking Conditions", J Acoust Soc of Amer, Vol. 55, S-20, 1974.
- [24] H. Hollien, R.E. McGlone, "An Evaluation of the 'Voiceprint' Technique of Speaker Verification",

1975 Carnahan Conference.

- [25] R. Geppert, M.H. Kuhn, H. Piotrowsky, H. Tomaschewski, "An Operational System for Personnel Authentication Using Long-Term Averaged Voice Spectrum", 1979 Carnahan Conference.
- [26] J.W. Bayless, S.J. Campanella, A.J. Goldberg, "A Survey of Speech Digitization Techniques", 1972 Carnahan Conference.
- [27] W. Haberman, A. Fejfar, "Automatic Identification of Personnel Through Speaker and Signature Verification System Description and Testing", 1976 Carnahan Conference.
- [28] A.D.C. Holden, Y.K. Galut, "A New Method for Accurate Analysis of Voiced Speech", IEEE Initial Conf on Acous Sp and Sig Proc, 1976.
- [29] A.D.C. Holden, J.Y. Cheung, Y.K. Galut, "The Role of Idiosyncracies in Linguistic Stressing Cues, and Accurate Formant Analysis in Speaker Identification, 1976 Carnahan Conference, pp.31-37.
- [30] L.L. Pfeifer, "Feature Analysis for Speaker Identification", Final Tech Report, RADC-TR-77-277, A044311. August 1977, Rome Air Dev Center, Griffiss AFB, NY.
- [31] J. Makhoul, "Linear Prediction: A Tutorial Review", Proc IEEE, Vol. 63, pp.561-80, April 1975.
- [32] J.D. Markel, A.H. Gray, "Linear Prediction of Speech", Springer-Verlag (1976).
- [33] B.T. Oshika, "FACP Speech Recognition/Transmission System", Final Tech Report, RADC-TR-78-193, August 1978, Rome Air Dev Center, Griffiss AFB, NY, A060115.
- [34] L.L. Pfeifer, "Inverse Filter Parameters for Speaker Identification", Final Tech Report, RADC-TR-76-71, June 1976, Rome Air Dev Center, Griffiss AFB, NY, A023823.  
also  
"Inverse Filter for Speaker Identification", Final Tech Report RADC-TR-74-214, August 1974, Rome Air Dev Center, Griffiss AFB, NY, 787860/6G1.

- [35] G.R. Doddington, B.M. Hydrick, "Speaker Verification II", Final Tech Report RADC-TR-75-274, November 1975, Rome Air Dev Center, Griffiss AFB, NY, A018901.
- [36] G.R. Doddington, B.M. Hydrick, R.E. Helms, "Speaker Verification III", Final Tech Report RADC-TR-76-262, August 1976, Rome Air Dev Center, Griffiss AFB, NY, B014720L.
- [37] R.L. Davis, B.M. Hydrick, G.R. Doddington, "Total Voice Speaker Recognition", Final Tech Report RADC-TR-78-260, January 1979, Rome Air Dev Center, Griffiss AFB, NY, A065160.
- [38] M.J. Foodman, "Test Results: Advanced Development Models of BISS Identity Verification Equipment, Volume II, Automatic Speaker Verification", MITRE Tech Report MR-3442, Vol. 11, September 1977, MITRE Corp., Bedford, Massachusetts.
- [39] M.B. Herscher, T.B. Martin, W.F. Meeker, "Automatic Speaker Identification and Verification", 1970 Carnahan Conference.
- [40] J.L. Flanagan, "Speech Analysis Synthesis and Perception", Springer-Verlag, 1965.
- [41] R. Wiggins, L. Brantingham, "Three Chip System Synthesizes Human Speech", Electronics, August 31, 1978, pp.106-116.
- [42] C.H. Coker, "A Model of Articulatory Dynamics and Control", Proc IEEE, Vol. 64, pp.452-60, April 1976.
- [43] B.S. Atal, "Automatic recognition of Speakers From Their Voice", Proc IEEE, Vol. 64, pp.460-75, April 1976.
- [44] C.A. McGonegal, L.R. Rabiner, B.J. McDermott, "Speaker Verification by Human Listeners Over Several Speech Transmission Systems", Bell Syst Tech J., Vol. 57, pp.2887-2900, October 1978.
- [45] B.S. Atal, L.R. Rabiner, "A Pattern-Recognition Approach to Voiced/Unvoiced/Silence Classification With Applications to Speech Recognition", IEEE Trans Acoust Speech & Sig Proc, Vol. ASSP-24, pp.201-12, June 1976.
- [46] S.S. McCandless, "An Algorithm for Automatic Formant Extraction Using Linear Prediction Spectra", IEEE Trans Acoust Speech & Sig Proc, Vol. ASSP-22, pp.135-41, April 1974.
- [47] W. Wolfram, R.W. Fasold, "The Study of Social Dialects in American English", Prentice Hall, NJ, 1974.

- [48] L. Herman, M.S. Herman, "American Dialects", Theatre Arts Books, NY., 1947.
- [49] H.L. Mencken, "The American Language",
- [50] A.J. Bronstein, "The Pronunciation of American English", Appleton-Century-Crofts, NY., 1960.
- [51] R.B. Mosen, A.M. Engebretson, "Study of Variations in the Male and Female Glottal Wave", J Acoust Soc of Amer, Vol. 62, pp.981-93, October 1977.
- [52] G.P. Krapp, "The English Language in America",
- [53] C.K. Thomas, "An Introduction to the Phonetics of the English Language", The Ronald Press Co., 2nd Ed., 1958.
- [54] H. Fletcher, "Speech and Hearing in Communication", Van Nostrand, Princeton, NJ, 1953.
- [55] R. Luchsinger, G.E. Arnold, "Voice-Speech-Language", trans. by: G.E. Arnold and E.R. Finkbeiner, Wadsworth Publ Co., 1965.
- [56] "Prevalence of Selected Impairments, U.S. - 1971", Data from National Health Survey, Series 1, No. 99, U.S. Report, HEW, 1975.
- [57] J.L. Flanagan, "Synthesis of Speech", Scientific American, Vol. 226, pp.48-58, February 1972.
- [58] 1970 United States Census, "Characteristics of the Population, Vol. 1; Part A-Number of Inhabitants, Section 1 and 2, part 1-United States Summary, Section 1; part 2-52 - Individual States; published by U.S. Dept of Commerce.
- [59] R.T. Gagnon, "Votrax Real-Time Hardware for Phoneme Synthesis of Speech", Proc IEEE Int Conf Acoust Speech & Sig Process, 1978.
- [60] M. Baumwolspiner, "Speech Generation Through Waveform Synthesis", Record, IEEE Int Conf Acoust Speech & Sig Process, 1978.
- [61] R.R. Leutnegger, "The Sounds of American English", Scott, Foresman & Co., 1963.
- [62] G. Fant, "Acoustic Theory of Speech Production", Mouton & Co., S-Gravenhage, 1960.

- [63] B.S. Atal, N. David, "On Synthesizing Natural Sounding Speech by Linear Prediction", Red, IEEE ICASSP, 1979.
- [64] J.E. Belyea, "MDEC Latent Fingerprint Recognition System", Carnahan Conference on Crime Countermeasures, pp.29-38, 1972.
- [65] P. Benson, "Test Results, Advanced Development Models of BISS Identify Verification Equipment, Vol. IV, Automatic Fingerprint Verification", MITRE Tech Report MTR-3442, Vol. IV, MITRE Corp., Bedford, Massachusetts, September 1977.
- [66] B.C. Bridges, "Practical Fingerprinting", Funk and Wagnalls, NY, 1942.
- [67] C.E. Chapel, "Fingerprinting - A Manual of Identification", Chapman & Hall, London, 1946.
- [68] R.P. Chiralo, L.L. Berdan, "Adaptive Digital Enhancement of Latent Fingerprints", Carnahan Conference, pp.131-36, 1978.
- [69] H. Cummins, C. Midlo, "Fingerprints, Palms and Soles", Dover Publications, Inc., NY, 1961.
- [70] M. Eleccion, "Automatic Fingerprint Identification", IEEE Spectrum, pp.36-45, September 1973.
- [71] S.B. Holt, "The Genetics of Dermal Ridges", Charles C. Thomas, Publ, Springfield, Illinois, 1968.
- [72] K. Millard, "An Approach to the Automatic Retrieval of Latent Fingerprints", Carnahan Conference, pp.45-51, 1975.
- [73] J. Mogilensky, "ADIT: Applying Technology to the Undocumented Alien Problem", Carnhan Conference, pp.137-42, 1978.
- [74] R.T. Moore, J.R. Park, "The Graphic Pen, An Economical Semi-Automatic Fingerprint Reader", Carnhan Conference, pp.59-62, 1977.
- [75] C.B. Shelman, "Fingerprint Classification - Theory and Application", Carnhan Conference, pp.131-38, 1976.
- [76] C.B. Shelman, D. Hodges, "Fingerprint Research at Argonne National Laboratory", Carnhan Conference, pp.108-13, 1973.

- [77] P.K. Shizume, C.G. Hefner, Jr., "A Computer Technical Fingerprint Search System", Carnahan Conference, pp.121-29, 1978.
- [78] R.B. Solosko, J.J. Paley, "A Semi-Automated Computer Fingerprint Encoding and Matching System", Carnahan Conference, pp.114-22, 1973.
- [79] R.M. Stock, "Automatic Fingerprint Reading", Carnahan Conference, pp.16-28, 1972.
- [80] C.E. Thomas, "Automatic Optical Fingerprint Identification", Carnahan Conference, pp.39-43, 1972.
- [81] G.L. Thomas, "Physical Methods of Fingerprint Development", Carnahan Conference, pp.38-51, 1975.
- [82] J.H. Wegstein, "Automated Fingerprint Identification", NBS Tech Note 532, National Bureau of Standards, U.S. Dept of Commerce, August 1972.
- [82] Z. Weinberger, J. Wasserman, C. Hillebrand, "Finident", Carnahan Conference, pp.117-24, 1976.
- [83] D.M. Osgard, S.A. Smithson, "Automated Fingerprint Identity Verification", Rockwell International, April 1977.
- [84] F. Victor, "Handwriting, A Personality Projection", Charles C. Thomas, Publ, Springfield, Illinois, 1952.
- [86] K.P. Zimmerman, C.L. Werner, "SIRYS - A Research Facility for Handwritten Signature Analysis", Carnahan Conference, pp.153-62, 1978.
- [87] J.E. LaRiviere, E. Simonson, "The Effect of Age and Occupation on Speed of Writing", J of Gerontology, pp.415-16, July 1965.
- [88] W. Kuckuck, B. Rieger, K. Steinke, "Automatic Writer Recognition", Carnahan Conference, pp.57-64, 1979.
- [89] U. Perret, "Computer Assisted Text Analysis", Carnahan Conference, pp.81-87, 1979.
- [90] S. Furui, "An Analysis of Long-Term Variation of Feature Parameters of Speech and its Application to Talker Recognition", Electronics and Communication in Japan, Vol. 57-A, #12, pp.34-42, 1974.

- [91] A. Fejfar, "Test Results, Advanced Development Models of BISS Identity Verification Equipment, Volume III, Automatic Handwriting Verification", MITRE Technical Report, MTR-3442, Vol. III, MITRE Corp., Bedford, Massachusetts, 1977.
- [92] E. Lee, "Universal Transparent Simulator", Final Technical Report, Contract Number, F30602-79-C-0226, Rome Air Development Center, Griffiss AFB, New York, October 1980.

APPENDIX A  
DATA CAT STATISTICS

A.1 NUMBER OF ATTEMPTS AND SUBJECTS FOR TYPE I ERROR TESTING

The purpose of this section is to answer a question posed by paragraph 4.1.1.1. of the Data CAT Statement of Work. "Determine the number of samples per individual and the number of separate sessions per individual required to determine a Type I error of .01 with a 90% and 95% confidence level."

We first assume there is no variability of the attribute or its measurement process from session to session. In this simplified case we will find the number of samples required in the data base. The question of how many sessions are required will be addressed in a later section.

Let the data base consist of  $N$  samples. An identity verification device is to be tested. The result is an acceptance or a rejection. Because of the assumption that there is no session-to-session variability, there is a single constant probability,  $p$ , that a sample will be rejected by the test. After all  $N$  samples are tested,  $M$  will have been found to be rejected. The problem is to estimate  $p$ , the

## DATA CAT STATISTICS

Type I error rate. The likelihood function for  $p$  is (Reference 1, p. 196):

$$L(p) = p^M (1-p)^{N-M}, \quad (A1)$$

and the most likely estimate of  $p$  is  $p^*$ , that value of  $p$  which maximizes the log  $L$ :

$$\frac{\partial \log L}{\partial p} = \frac{M}{p} - \frac{N-M}{1-p} = 0 \quad (A2)$$

Solving Equation A2 yields

$$p^* = \frac{M}{N} \quad (A3)$$

which is, of course, the intuitive estimate for the Type I error as well.

A confidence interval about  $p^*$  is defined by a single parameter  $\Delta p$ , which is said to provide a confidence level of  $C$  when

$$C = \frac{\int_{p^* - \Delta p}^{p^* + \Delta p} Ldp}{\int_0^1 Ldp} \quad (A4)$$

Equation A4 means that the statement "The Type I error has a value

## DATA CAT STATISTICS

$p^* \pm \Delta p$ ." will be true 100 percent of the time.

Because of the shape of  $L$ , for certain values of  $M$  and  $N$ , it will not be possible to find a single  $\Delta p$  which provides the desired confidence and at the same time keeps both  $p^* + \Delta p$  and  $p^* - \Delta p$  within the known range of  $p$  (from 0 to 1). In this case, a logical interpretation of Equation A4 is to replace  $p^* + \Delta p$  with 1 or  $p^* - \Delta p$  with zero, depending upon which limit was exceeded.

We would now like to plot  $L$  for a reasonable value of  $M$  and  $N$  in order to gain some insight into its behavior. What is a typical value of  $M$ ? This is turning the problem about the other direction. Previously we have been considering a best estimate for  $p$  given  $M$ . Now we want to know a typical  $M$ , which, of course, can only be answered by knowing  $p$ . The value of  $p$  of interest for this study is .01. Thus, we now ask for the most likely value of  $M$  given  $p$ . Clearly

$$M^* = pN \tag{A5}$$

and, in fact, the probability that any value of  $M$  will be observed is given by the binomial distribution

$$\pi(M) = \frac{N!}{M!(N-M)!} p^M (1-p)^{N-M} \tag{A6}$$

Taking  $N=100$  samples, we find a most likely value of  $M$  to be 1. A plot of  $L(p)$  for  $N=100$  and  $M=1$  is shown in Figure A1. Note that if a confidence of .95 were specified, the interval about  $p^*$  would be asymmetric. The lower value of  $p$  would be zero while the upper would be well above .03.

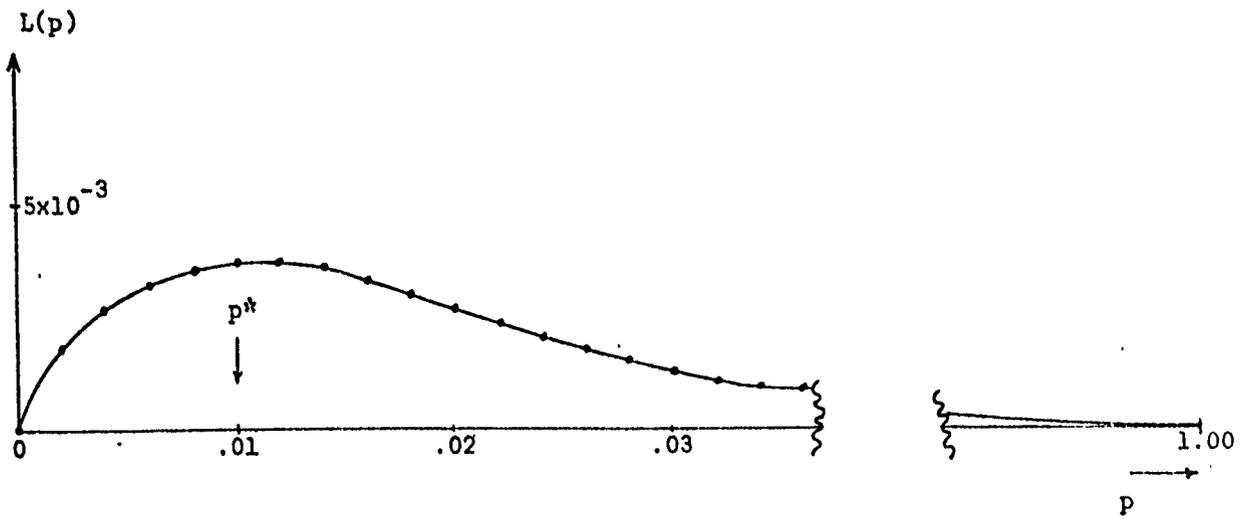


Figure A1 The Likelihood Function for  $N=100$ ,  $M=1$

## DATA CAT STATISTICS

The above discussion should make it clear that confidences are defined on intervals of  $p$ . The problem before us is to find a sample size which will provide testing to a specified confidence for  $p=.01$ . This is insufficient information, as an interval about  $.01$  must also be set. What is a reasonable value for the interval? That is, what accuracy on  $p$  is desired? We submit that if the value of  $p$  required on the identity verification system is  $.01$ , then one would like to distinguish between the case  $p = .01$  and  $p = .02$ . (One is not really interested in knowing that  $p = .00981 \pm .00001$ , for example, although given sufficient samples this level of accuracy could be achieved.) Thus, a reasonable value of  $\Delta p$  for  $p = .01$  is  $\Delta p = .005$ . This will permit the 1% and 2% Type I error cases to be distinguished.

The question which we have set out to answer may now be posed. "What sample size  $N$  is required to permit  $p$  in the neighborhood of  $.01$  to be determined to  $\pm .005$  with a confidence of 90% or 95%." It is clear from Figure A1 that  $n=100$  samples is insufficient. As  $N$  grows larger, with  $N=pN=.01N$  fixed,  $L$  approaches a Gaussian shape,

$$L(p) \rightarrow \exp \left( -\left(p - \frac{M}{N}\right)^2 / 2M\left(1 - \frac{M}{N}\right) \right) \quad (A7)$$

Equation A7 is easily derived by expanding  $\log L$  in a Taylor series about  $p^*$ . From the normal curve of error 90% of the area is contained within 1.645 of the standard deviation and 95% within 1.96. Thus, we require a value of  $N$  such that

$$1.645 \sqrt{M\left(1 - \frac{M}{N}\right)} \leq \frac{1}{2} (.01)N \quad (A8)$$

## DATA CAT STATISTICS

Inserting  $h = .01N$  gives

$$1.645 \sqrt{(.01)N(.99)} \leq .005 N \quad (A9)$$

$$1071.6 \leq N \quad (A10)$$

The 95% confidence value is  $1521.3 \leq N$

The estimates based on assuming a Gaussian shape can be refined to exact answers by performing the integration of Equation A4. The integrals can be written as:

$$C = [B_{p^*+\Delta p}^{(M+1, N-M+1)} - B_{p^*-\Delta p}^{(M+1, N-M+1)}] / B^{(M+1, N-M+1)} \quad (A11)$$

Where  $B_x$  and  $B$  are the incomplete and complete beta functions [2, p.263]. Using an expansion for  $B_x$  good for small, non-zero values of  $x$  [2, p.944], we can compute a confidence table, Table A1 for  $p = .01$ . Table A1 is used by finding a confidence interval of interest in the top row and a number of samples in the left hand column. The intersection of row and column gives the confidence value. We have recommended an interval of  $\pm .005$  about  $p = .01$ . This is tabulated in the second column. From Table A1 we see that 1200 samples would produce a 90% confidence on this interval and 1800 would produce 95% confidence.

The above discussion has made no mention of the number of different individuals included in the study. It simply says that a binary decision making device must be tested 1200 times to establish that  $\Delta p = \pm .005$  when  $p = .01$ . Suppose now that the access system is tested on two people. The requirement that the system perform at  $p = .01$  can be interpreted in two different ways. The Type I error

TABLE A1

Confidence Values for Selected Number of Samples  
and Intervals About  $p = .01$  for the Binomial Distribution

# Samples	p Interval				
	[.0075, .0125]	[.005, .015]	[0, .02]	[0, .03]	[0, .04]
100		18.3	35.4	59.7	85.3 90.7
200		26.2	50.1	76.2	94.2 98.8
500		41.9	72.0	93.5	99.7
1000		56.8	87.0	99.0	
1100			88.7	99.3	
1200		61.1	90.2	99.4	
1500		66.5	93.4	99.8	
1600		68.0	94.2	99.9	
1700		69.5	94.8		
1800		70.9	95.4		

## DATA CAT STATISTICS

could be the average system performance or it could be the performance requirement on each individual. That is, if Individual 1 is tested 10,000 and rejected 175 and Individual 2 is rejected 25 times out of 10,000, then the system performance is  $p = .01$  in the average sense even though Individual 2 has shown a  $p = .0175$ .

Which of the two interpretations above makes the best sense for testing an access device? It seems obvious that in a large human population one will always be able to find a subject whose measurements are sufficiently variable to reproduce a  $p$  greater than .01. For an acceptable access control system, however, the number of such subjects should be vanishingly small. Thus, the interpretation of a  $p$  specification as a system average is the sensible one. It follows that the number of samples we have computed is the total samples for all individuals, not the number of samples per individual.

The foregoing argument would make it appear that 1200 samples could be drawn from 1200 subjects, one sample per subject (in addition to the samples needed for enrollment). We would now like to show that there is a more realistic lower bound on the number of samples per subject.

What will determine the number of subjects in the study? First of all, we note that fewer subjects in the system permits economy of data collection and storage because a fixed number of samples per subject must be collected for enrollment. Say 100 enrollment samples are collected for 1200 subjects. The enrollment data base is 120,000

## DATA CAT STATISTICS

samples while the total data base of enrollment and test data bases is only 1200 greater, continuing the example of one test sample per individual. At the opposite end of the scale, if only one individual is to be used for the data base with 1200 test samples for him, then the total data base consists of only 1300 samples. It is also apparent that the collection of data from one individual would be easier and less costly than from 1200.

Despite the foregoing, it is obvious that the data base must include more than one subject because the population of subjects will not be homogeneous with respect to the attribute being measured. In collecting the signature data base, for example, we know a priori that there are two fundamental groups of subjects, left- and right-handed persons. Subjects must be drawn from all groups of a significant size for which there is reasonable probability of systematic attribute variation. Let us suppose that 20% of the population is left-handed and 80% right-handed. Then a possible procedure is to use one subject from each of the two classes and to collect four times more samples from the right-hander. Since the average  $p$  will be computed as the weighted sum of the right-handed Type I error,  $p_R$ , and the left,  $p_L$ ,

$$p = .2 p_L + .8 p_R \quad (A12)$$

error in  $p_R$  contributes more to error in  $p$ . An alternate and superior procedure is to use one left-handed subject and four right. Then an equal number of samples should be collected from each.

## DATA CAT STATISTICS

That subgroups are expected in the population demonstrates that previously undiscovered subgroups may be revealed in the testing of a new access device. This is an additional reason why there must be a number of different subjects in the data base. For example, suppose the data base consisted of samples from two randomly selected subjects and there are two undiscovered subgroups, each comprising 50% of the population. Further suppose they have a Type I error of 1.9 and .1%. Two times out of four the individuals in the data base would include a subject from each subgroup, one time in four it would include two subjects from the first subgroup, and one time in four, two from the second. Assume enough samples per subject that the Type I error for the first subject,  $p_1$ , and for the second,  $p_2$ , are known with high precision. Then the true  $p$  for this population,  $p_{\text{true}}$ , is 1.0%. However, because of too few subjects in the data base, a value of  $p$  different from  $p_{\text{true}}$  can result. This is shown in Table A2, where the three cases are given at the left of the Table, each with its probability of occurrence. The value of  $p$  which would be computed is shown in the column labelled 'p', and the square deviation from  $p_{\text{true}}$  in the last column. The rms deviation is .64%.

We now consider the same situation more generally. Instead of two subgroups with discrete value of  $p$ , we permit a continuum of possible values of  $p$ . Now let  $f(p)dp$  be the fraction of the population having Type I error,  $p$ , between  $p$  and  $p+dp$ . We want to know how many subjects to include in our sample in order to prevent a widely spread distribution  $f$  from affecting the results. Again,

TABLE A2

The Error in Type I Error Estimate  
Caused by Having Two Subgroups

Subgroup	Probability	$p = 1/2(p_1 + p_2)$	$(p - p_{TRUE})$
1	2		
x	x	0.5	0
xx		0.25	.000081
	xx	0.25	.000081

$$\overline{(p - p_{TRUE})^2} = .0000405$$

$$\Delta p = \sqrt{\overline{(p - p_{TRUE})^2}} = .0064$$

## DATA CAT STATISTICS

assuming that enough samples are taken from each subject so that the value of  $p$  for the subject may be determined with ignorable error, we can state that the average value of  $p$  given by

$$\bar{p} = \int_0^1 p f(p) dp \quad (A13)$$

is the value of  $p$  for the whole population and is therefore the value we would want our sample to represent. Thus, we need to have enough subjects so that  $\bar{p}$  is determined with small error. The accuracy of an estimate of  $\bar{p}$  is also determined by the variance of the distribution  $f$ . In fact, from the Central Limit Theorem we can state that the error in a determination of  $\bar{p}$ ,  $\sigma$ , is given by

$$\sigma = \frac{\Delta p}{\sqrt{k}} \quad (A14)$$

as  $K$  grows large. Here  $\Delta p$  is the standard deviation of  $p$  due to the distribution  $f$ ,

$$\Delta p^2 = \int_0^1 (p - \bar{p})^2 f(p) dp \quad (A15)$$

Actually, the type of distribution which produces the largest  $\Delta p$ , and, hence, according to Equation A14 the largest  $\sigma$  is a binomial distribution of the sort we considered in the example of Table A2. We will make this worst case assumption in order to establish an upper bound on  $K$ . Let  $f_1$  be the fraction of the population belonging to Subgroup 1. Let  $p_1$  be the Type I error of this subgroup. Let  $f_2$  and  $p_2$  be the corresponding quantities for Subgroup 2.  $K$  subjects are selected randomly. The true value of  $p$

## DATA CAT STATISTICS

for the population is:

$$\bar{p} = f_1 p_1 + f_2 p_2 \quad (\text{A16})$$

A particular draw of K subjects will consist of v members of Subgroup 1 and K-v of 2. The probability,  $\pi$ , of this event is

$$\pi(v) = \binom{K}{v} f_1^v f_2^{K-v} \quad (\text{A17})$$

The resulting Type I error which would be measured is

$$p = p_2 + \frac{v}{K} (p_1 - p_2) \quad (\text{A18})$$

Now,

$$\sigma^2 = \sum_{v=0}^K \pi(v) (p - \bar{p})^2 \quad (\text{A19})$$

which gives

$$\sigma = |p_1 - p_2| \sqrt{\frac{f_1(1-f_1)}{K}} \quad (\text{A20})$$

Fixing  $\bar{p}$  at .01, from Equation A16

$$p_2 = \frac{.01 - f_1 p_1}{1 - f_1}, \quad (\text{A21})$$

giving

$$\sigma = |.01 - p_1| \sqrt{\frac{f_1}{1-f_1}} \frac{1}{\sqrt{K}} \quad (\text{A22})$$

The worst case value (large  $\sigma$ ) is produced by  $f_1$  near 1. Since

## DATA CAT STATISTICS

$p_2$  is less than or equal to one, from A21

$$f_1 \leq \frac{.99}{1-p_1} \tag{A23}$$

Thus, the worst case value of  $f_1$  is for  $p_1 = 0$  and  $f_1 = .99$ .

Inserting these values into A22 gives

$$\sigma \leq .01 \sqrt{\frac{.99}{k}} \tag{A24}$$

Using again the requirement that the sample size should be sufficient to distinguish a Type I error of .01 from that of .02,

$$\sigma \leq .01 \sqrt{\frac{.99}{k}} \leq .005 \tag{A25}$$

yields

$$396 \leq k \tag{A26}$$

Table A3 shows the probabilities of measuring certain  $\bar{p}$  values when

400 subjects are used and the worst case assumptions are made. Observe that a value of  $\bar{p} = 1.50\%$  or less is obtained 29% of the time.

Unfortunately, the result  $K=400$  is rather a large number of subjects to include in the data base. This large number has arisen due to the fact that we have postulated a subgroup comprising only 1% of the population. If we were to relax the specifications so that only subgroups of 5% or more would be of concern, then  $f_1$  can be set

TABLE A3

Two Subgroups Assumed, With Type I Errors 0.0 and 1.0  
 And Frequency of Occurrence .99 and .01.

Table shows probability of occurrence for  
 various  $\bar{p}$  values for 400 subjects in sample.

$v$	$\pi(v)$	$\bar{p}$
400	.01795	0
399	.07253	.25%
398	.14615	.50%
397	.19585	.75%
396	.19635	1.00%
395	.15708	1.25%
394	.10446	1.50%
393	.05939	1.75%

## DATA CAT STATISTICS

to .95. Using again the worst case values of  $p_1 = 0.0$  and  $p_2 = 0.2$ ,

$$.005 = \sigma \leq .2 \sqrt{\frac{(.5)(.95)}{k}} \quad (\text{A27})$$

or

$$28 \leq k \quad (\text{A28})$$

Similarly, if subgroups no smaller than 10% of the population are considered

$$8 \leq k \quad (\text{A29})$$

## DATA CAT STATISTICS

### A.2 NUMBER OF ATTEMPTS AND SUBJECTS FOR TYPE II ERROR TESTING

The purpose of this Section is to discuss the number of attempts which must be made against an identity verification device to ascertain its Type II error performance to certain confidence levels.

We first define the Type II error rate. Let  $\alpha$  be an index over the population to be enrolled in the system. An individual who is enrolled will have an 'account' which will contain the personal data against which he will be compared. When another individual,  $i$ , makes a verification attempt against account  $\alpha$ , an opportunity for a Type II error arises. The probability that individual  $i$  will be accepted under account  $\alpha$  will be denoted  $p_{\alpha i}$ . By letting  $i$  run over all members of the population which might attempt access, we could obtain the Type II error rate of account  $\alpha$ ,

$$P_{\alpha} = \frac{1}{N_{TOT}} \sum_{i=1}^{N_{TOT}} P_{\alpha i} \quad , \quad (A30)$$

where  $N_{TOT}$  is the size of the intruder population.

In Section A.1 we discussed for Type I errors whether a specification on the error rate should be a rigid bound on all accounts or an average over all accounts. We demonstrated that only the latter made sense. Correspondingly, we here adopt a definition of the identity system Type II performance as an average. The Type II error rate is defined as

$$P = \frac{1}{N_{TOT}^1} \sum_{\alpha=1}^{N_{TOT}^1} P_{\alpha} \quad , \quad (A31)$$

## DATA CAT STATISTICS

where  $N'$  is the size of the population of individuals who would potentially be enrolled in the system.

Since the goal of Data CAT is to collect a general data base, neither intruder nor enrollee population can be specified exactly. We take both to be the entire American population. Thus,

$$p = \frac{1}{N_{TOT}} \sum_{\beta} \sum_{\alpha}^{N_{TOT}} p_{\alpha\beta}, \quad \alpha \neq \beta \quad (A32)$$

where the two populations are considered to be the same and an individual is eliminated from the Type II statistics against his own account by deleting the  $\alpha = \beta$  term.

Altogether, two random variables must be adequately sampled in compiling the Data CAT data base. There should be enough access attempts that the individual terms  $p_{\alpha\beta}$  are accurately estimated, and there should be sufficient account-intruder pairs that the population is adequately sampled.

Despite the foregoing, to simplify the discussion we first assume all accounts and intruders are equivalent. We have a single account and a single intruder. We want to know how many samples,  $N$ , of the intruder are required to test a system which performs with a Type II error near a) .02 and b) .00001. The intruder is either accepted or rejected so the binary statistics developed in Section 1.0 can be used.

## DATA CAT STATISTICS

Table A4 gives the confidence for a system with  $p = .02$ . For example, if the data base contained 200 samples and four were falsely accepted, then the Type II error would be 2%. Furthermore, by examining the Table, one sees that the assertion that  $p = .02 \pm .01$  has a 67% confidence. Using this error interval as the most reasonable choice, we can state that 800 samples are required for 90% confidence and 1000 for 95%. Table A5 provides the same information as Table A4 for a system with  $p = .001\%$ . Here we see that for the preferred choice of  $p = .001\% \pm .0005\%$ ,  $1.2 \times 10^6$  samples are sufficient for 90% confidence, but even  $1.5 \times 10^6$  samples are insufficient to achieve a confidence of 95%. Comparing to Table A4 we estimate a requirement of  $1.8 \times 10^6$  samples.

We now extend the argument to consider the fact that different account-intruder pairs will have different values of  $p$ . We must have a sufficient number of pairs to sample the population adequately. This question was also considered in Section 1.C under the "undisclosed subgroup problem." There we showed that the greatest danger of biased sampling occurred for two subgroups, one with  $p_1 = 0$  comprising 98% of the population and one with  $p_2 = 1.00$  comprising 2%. This produces a  $p$  equal to .02 but a large sample of individuals is required to reduce the fluctuations in the number of members of the poorly performing subgroup included in the sample. From Equations A16 and A22,

$$\sigma = .02 \sqrt{\frac{.98}{1-.98}} \frac{1}{\sqrt{k}} \leq .01 \quad (A33)$$

TABLE A4

Confidence Levels for a System  
With Type II Near 2%

Interval $p = .02 \pm$	Interval		
	.005	.01	.02
N Samples			
100		27.0	50.5 77.0
200		38.0	67.0 90.7
500		57.1	87.3 99.0
800		68.3	91.1 99.9
1000		73.7	96.5 99.95
1500		82.9	98.9 99.98

TABLE A5

Confidence Levels for a System  
With Type II Near .001%

Interval $p = 10^{-5} \pm$	$.25 \times 10^{-5}$	$.5 \times 10^{-5}$	$10^{-5}$
N Samples			
$10^5$	18.2	35.2	59.4
$10^6$	56.5	86.8	98.9
$1.2 \times 10^6$	60.8	89.9	99.4
$1.5 \times 10^6$	66.2	93.2	99.8

## DATA CAT STATISTICS

Thus, the number of pairs required,  $K$ , is

$$196 \leq k \tag{A34}$$

For the device performing at  $p = .001\%$ , 399996 pairs would be required.

Notice that the population of intruders and accounts cannot overlap. However, a population of enrolled individuals must be collected for Type I testing, anyway. The intruders could be drawn from a subset of this group. That is, suppose  $K'$  individuals are collected for Type I error testing. Let their accounts be numbered 1, 2, ...,  $K'$ . Use Account 1 and run individual 2, 3, ...,  $K'$  as intruders. Use Account 2 and run the  $K'-2$  remaining individuals as intruders. Notice that Individual 2 is run against Account 1 but not conversely. Choosing both combinations would not constitute an independent sample from the universe of all possible pairs even though  $p_{\alpha\beta}$  is not necessarily equal to  $p_{\beta\alpha}$ . Proceeding in this fashion, one obtains  $(K'-1)K'/2$  pairs. Thus, with  $K'$  individuals in the data base, the number of tests which may be performed,  $K$ , is

$$k \cong \frac{k'^2}{2} \tag{A35}$$

## DATA CAT STATISTICS.

For Type II error of 2%, using Equations A34 and A35, approximately 20 individuals are required. For an error of .001%, approximately 895 individuals are required.

Just as the total number of pairs is a quadratic function of the number of individuals, the total number of samples required,  $N$ , is a quadratic function of the number of samples per individual,  $n$ . Thus,

$$N = \frac{(k'n)^2}{2} \quad (A36)$$

For example, we know that 399996 ( $=K'^2/2$ ) pairs are required for  $p = .001\%$ . Also,  $1.2 \times 10^6$  ( $=N$ ) tests are required to establish a 90% confidence on the recommended interval  $\pm .0005\%$ . Thus, inserting into Equation A36,

$$n = 3 \quad (A37)$$

Table A6 summarizes the requirements on individuals and samples in the data base. We observe that a requirement for 400 individuals to achieve a Type I error rate of 1% is more demanding than the equivalent twenty individuals for a 2% type II.

TABLE A6

Type II Error	Number of Persons K'	Samples/Person	
		90%	95%
2%	20	4	5
.001%	895	3	5

## DATA CAT STATISTICS

### A.3 NUMBER OF SAMPLES FOR ENROLLMENT

Currently available identity verification systems measure a personal attribute of a test subject and compare the measured attribute with a previously stored reference file for the subject. This operation, the verification process, thus requires a reference file for each user. The reference file is created when the user is enrolled in the system, but may be updated with subsequent measurements from verification attempts.

The purpose of Data CAT is to design a data base of speech, fingerprint, and handwriting attributes which will permit testing of potential identity verification devices. The data base must contain measurements for both enrollment and verification. Typically, at enrollment several repeated measurements are performed. For example, a subject in the handwriting system would sign his name several times in order to establish a representative pattern. The Data CAT data base should be general enough to accomodate a wide class of verification systems, and, therefore, the enrollment portion must contain more than one measurement of the attribute. Each measurement will be called a sample. This Section will discuss the number of samples which should be collected for the enrollment portion of the data base.

## DATA CAT STATISTICS

All identity verification devices currently under consideration by the Air Force work in a fashion which is easily described in the language of linear pattern recognition. When the personal attribute is collected, certain key features, believed particularly individual and stable, are extracted. The complex attribute is thus reduced to a simpler set of features. Let us suppose that  $\Lambda$  measurements are made, the first measurement having value  $x_1$ , etc. The measurements, assembled as a vector,  $x$ ,

$$x = \langle x_1, x_2, \dots, x_\Lambda \rangle, \quad (A38)$$

comprise the feature vector. In linear pattern recognition the vector  $x$  is treated as a point in a  $\Lambda$ -dimensional linear space. When the attribute for the subject is measured a second time, due to measurement noise, statistical fluctuation, or actual change in value, the feature vector,  $x$ , will be different. However, if the personal attribute is useful for identification and the features are well constructed, then all the vectors for a particular subject should be relatively close together and relatively far from vectors belonging to a different subject. A metric is obviously needed to formalize the notion of distance. Figure A2 shows a two-dimensional space with feature vectors for two subjects.

At enrollment a reference file for a subject is created. This reference file is a means of specifying that region or regions in feature space which are likely to contain vectors for the subject. At verification a newly acquired feature vector is tested to see whether it lies in an acceptable region for the subject, and he is accepted or

Measurement 2

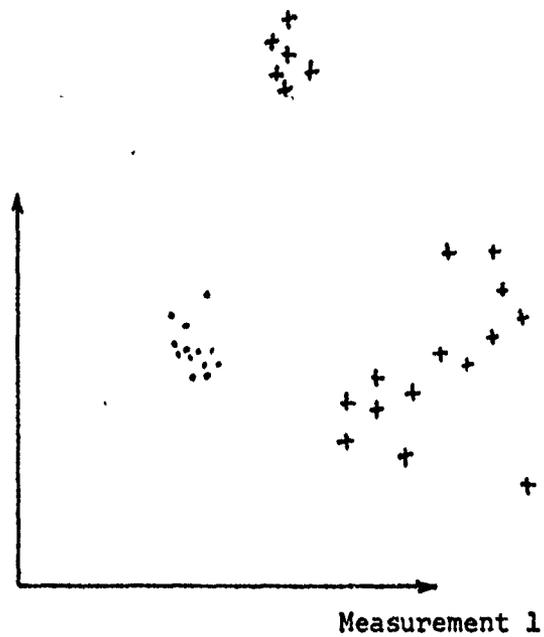


Figure A2 A two-dimensional feature space containing measurement vectors for two subjects, one represented by dots, the other by crosses.

## DATA CAT STATISTICS

rejected accordingly.

A number of methods for specifying the so-called decision boundaries of regions are in common use. In general, there are simple methods which involve few parameters and represent the regions by relatively simple shapes and methods employing numerous parameters, thereby capable of representing more complex shapes. Each parameter which is used to specify a region must be established before a decision strategy can be implemented. In an identity verification system the region parameters are estimated at enrollment by using the repeated measurements of the attribute under consideration. The accuracy with which a region can be specified will depend both on the number of parameters needed and on the dimensionality of the space. Moreover, the number of samples of an attribute which are available to estimate the parameters directly affects the accuracy of region representation.

For Data CAT we need to determine the number of samples of an attribute which might be required by a future identification device. The answer to this question depends on the dimensionality of the feature space and on the complexity of the region, both of which is impossible to describe without previously specifying the device. We can, of course, make estimates of the maximum dimensionality permitted in the data for the respective attributes. Furthermore, we could postulate commonly used region shapes (or decision strategies). We postpone this ultimate question for the present and consider a few

## DATA CAT STATISTICS

well known decision strategies in order to elucidate the interplay between complexity of a strategy and number of samples required.

The simplest strategies, or 'logics' as they are sometimes called to emphasize their decision making role, all assume a single simply connected region. A straightforward logic is to assume that the regions for all individuals may be represented by a simple geometric figure such as a hypersphere, hypercube, or hyper-rectangle. The size and shape of the geometric figure are fixed, only its location need be ascertained by samples of feature vectors for the subject. Figure A3 shows such a decision strategy for three individuals.

How many samples are required to center the decision box? Suppose for the present that  $\Lambda = 1$  and  $n$  measurements are made:  $x^{(1)}$ ,  $x^{(2)}$ , ...,  $x^{(n)}$ . Then the average value of  $x$ , where the box should be located is:

$$\bar{x} = \frac{1}{n} \sum_{\alpha} x^{(\alpha)} \quad (\text{A39})$$

The best estimate for the error in each measurement is

$$\Delta x = \sqrt{\frac{1}{n-1} \sum_{\alpha} (x^{(\alpha)} - \bar{x})^2} \quad (\text{A40})$$

and the best estimate of the deviation of  $\bar{x}$  from the true mean is

$$\Delta \bar{x} = \frac{\Delta x}{\sqrt{n}} \quad (\text{A41})$$

where the estimate of the mean given by Equation A39 will lie within  $\pm \Delta \bar{x}$  of the true mean with probability .682. The observed  $x$  will be required to be smaller than some bound  $B$  (presumably related to the

Measurement 2

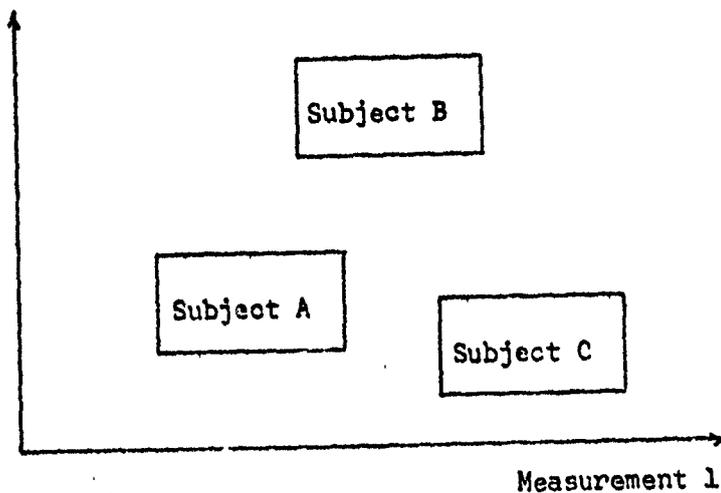


Figure A3 Simple decision logic utilizing fixed geometric shapes. A measurement vector lying inside box A is accepted as Subject A, etc.

DATA CAT STATISTICS :

characteristic size of the box) in order to make the estimated mean,  $\bar{x}$ , close to the true mean,

$$\bar{x} \leq B \tag{A42}$$

Equations A40, A41, and A42 give an operational test to determine whether enough measurements have been made. Now suppose  $x$  is two-dimensional. Then the mean and standard deviation for each component are calculated as above. But to guarantee that both components have a standard deviation close to the true mean will require more measurements. Suppose we require that  $\bar{x}_1$  and  $\bar{x}_2$  lie within  $B_1$  and  $B_2$  of their true mean with probability .682. Then we must require that  $\bar{x}_1$  and  $\bar{x}_2$  lie within the bounds with probability .826. In general, each component must satisfy its bounds with a probability  $.682^{1/\Lambda}$  to make the joint probability .682. For example, if we make sufficient measurements that  $\Delta\bar{x}/B \leq .5$ , then  $x$  will lie within  $B$  of the true value with probability .954, since

$$.954 = 2 \left( \frac{1}{2\sqrt{\pi}} \int_0^2 \exp\left(-\frac{y^2}{2}\right) dy = \operatorname{erf} \frac{2}{\sqrt{2}} \right) , \tag{A43}$$

or in general,

$$(.682)^{1/\Lambda} = \operatorname{erf} \left( \frac{B}{\sqrt{2} \Delta\bar{x}} \right) . \tag{A44}$$

Substituting Equation A40 gives

$$(.682)^{1/\Lambda} = \operatorname{erf} \left( \frac{B \sqrt{n}}{\sqrt{2} \Delta x} \right) . \tag{A45}$$

## DATA CAT STATISTICS

The value of  $n$  in units of  $(B/\Delta x)^2$  for some representative choices of  $\Lambda$  are given in Table A7.

Another method for specifying a region is to estimate not only the location of a simple geometric figure, but also its size. Figure A4 shows such a use of ellipses. For each ellipse the location and width of the ellipse must be determined by sampling, for a total of  $2\Lambda$  parameters per class. Using Equation A45 with  $\Lambda$  replaced by  $2\Lambda$  produces Table A8.

Another common method of specifying a logic is to permit the geometric figures to have arbitrary size and orientation in addition to location. In this case  $\Lambda^2$  parameters are used for size and orientation for  $\Lambda$  and location. Using Equation A45 with  $\Lambda$  replaced by  $\Lambda^2 + \Lambda$  produces Table A9.

Another condition which may be placed on the number of samples required is that the estimated parameters be linearly independent. This can occur only if the number of samples numbers is greater than the number of estimated numbers. With  $n$  samples of dimension  $\Lambda$ ,  $n\Lambda$  numbers are available. Thus, at least one sample is required for a  $\Lambda$  parameter logic, two are required for a  $2\Lambda$  logic, and  $\Lambda + 1$  are required for a  $\Lambda^2 + \Lambda$  logic.

In conclusion, we observe that the number of samples required to establish the parameters of a region is dependent on the type of logic employed and on the dimensionality of the feature space. However, for logics of the first two types considered, in which the number of parameters to be estimated is a linear function of  $\Lambda$ , even for large

TABLE A7

Logic With  $\Lambda$  Parameters Per Class

$\Lambda$	1	10	100	1000
$(\frac{B}{\Delta x})^2 n$	1	4.28	8.32	12.7

TABLE A8

Logic With  $2\Lambda$  Parameters Per Class

$\Lambda$	1	10	100	1000
$(\frac{B}{\Delta x})^2 n$	1	1.84	8.71	17.1

TABLE A9

Logic With  $\Lambda^2 + \Lambda$  Parameters Per Class

$\Lambda$	1	10	100
$(\frac{B}{\Delta x})^2 n$	1.84	8.71	17.1

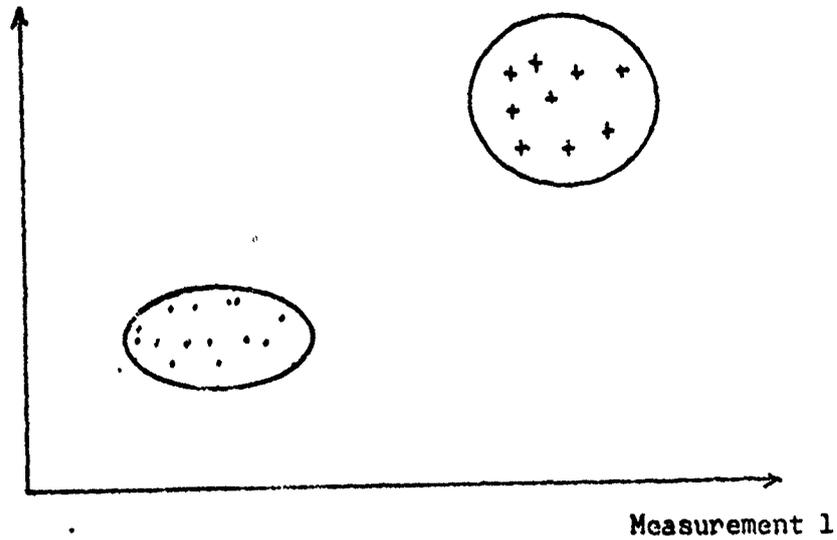


Figure A4 Ellipses of adjustable dimensions used to specify regions.

## DATA CAT STATISTICS

$\Lambda$  the number of measurements is not that excessive.

We now turn to a consideration of the maximum dimensionality available in the attributes considered. In the case of speaker identification, we shall presume that an utterance of two seconds is used consisting of approximately twenty different phonemes [3]. Each phoneme can be characterized by a few numbers such as voiced/unvoiced, pitch, and formant position, bandwidth, and relative amplitude. Altogether, some twenty numbers are perhaps sufficient, leading to an estimate of  $\Lambda \approx 400$  for a two second utterance. It is interesting to compare this to the number of bits necessary to encode the utterance. Using either a channel or LPC vocoder, approximately 2000 bits/second are required for good quality speech [4].

Whereas voices are adequately decomposed into formants, no such set of features has even been devised for fingerprints. Much of the information content of a print resides in the minutiae, however. Assuming four numbers per minutia (two for location, one for direction, and one for type) and 100 minutiae per print yield an estimate of  $\Lambda \approx 4000$  for all ten fingers. Encoding of fingerprints requires some 60,000 bits per digit [5].

In the case of signatures, even less is known, and neither estimates of the number of features nor the number of bits for encoding are available in the literature. Assuming an average

## DATA CAT STATISTICS

signature to be perhaps five inches in length when integrated along its arc and assuming a resolution of .01 inches, then the spatial information might require some 1000 numbers. At 10 bits per number, 10000 bits per signature would be required. Finally, doubling this number to allow for a pressure variable yields 20000 bits. This number is an upper bound since many portions of a signature consist of line segments of low or zero curvature.

Considering the current speaker verification system built by Texas Instruments [6] to be prototypical, we can compare the number of dimensions utilized and the number of enrollment samples required. In each utterance four reference points with 100 associated numbers are evaluated, giving a dimensionality of 400. Since these reference points concern only vowels, not all phonemes are exploited. Thus the agreement between the theoretical and actual dimensionality is largely coincidental. At enrollment time, each word is spoken four times.

We consider, likewise, the fingerprint verification device built by CALSPAN Corporation [7] to be typical. Unfortunately, the operation of the device is not described in open literature. Although print matching is based on minutiae (position in two coordinates and orientation), the number of minutiae used is not stated. It appears to be variable depending on the number located within the print, with three being a minimum. Thus, the dimensionality is greater than nine. The CALSPAN device requires ten enrollment samples.

## DATA CAT STATISTICS

Even less information is available about the operation of the prototypical signature verification system built by Veripen [8]. Veripen uses six signatures for enrollment.

Table A10 summarizes this information. The conclusion which can be reached from Table A10 is that the number of enrollment samples is consistent with the dimensionality presuming a simple logic is employed. This is known to be the case for the Texas Instruments device which uses the following simple region specification. The logic employed is to require that the measured vector  $x$  lies within a distance  $t$  of the reference vector  $r$ . That is

$$\sum_{\alpha} (x_{\alpha} - r)^2 < t \quad (\text{A46})$$

The variable  $t$  is allowed to be a function of individual. Equation A46 thus defines a circle of variable radius in feature space and is a very simple example of the second type of logic which we discussed.

Based on Tables A7, A8, A9, and A10, it would appear that ten samples taken at enrollment is a reasonable number. A rather compelling argument for using such a small number is the observation that no practical access control device can require too many enrollment samples. If this were the case it would be unacceptable to both users and agencies deploying it. As a conservative measure to guard against possibly unusable data, we recommend that the minimum number of enrollment samples be doubled to twenty.

TABLE A10

THEORETICAL AND ACTUAL DIMENSIONALITY  
OF FEATURE SPACE FOR THREE ATTRIBUTES

Attribute	Device	Bits	Theoretical Dimensionality	Used Dimensionality	Enrollment Samples	
Speech		TI	4000	400	400*	4
Fingerprint	Calspan	50000		≥400	≥9	10
Signature	Veripen	≤10000		--	--	6

\* Probably not linearly independent.

## DATA CAT STATISTICS

### A.4 NUMBER OF SESSIONS FOR EACH SUBJECT

As we have shown in Section 1.0,  $N$  samples are required for Type I error testing and  $K$  subjects must be included. Thus, at least  $N' = N/K$  samples per subject are required in the data base. This note is concerned with the question of how many sessions should be used to collect the  $N'$  samples.

As a basic premise we assume that as the number of sessions increases, the cost of collection will go up. This is reasonable since in any data collection there are the overhead expenses of set-up time, travel time, subject coordination, and general organization. In fact, it is usually the case that the time devoted to overhead items dominates the total time allocated. Therefore, we should minimize the number of separate sessions.

It will not ordinarily be possible, however, to collect all the required data in a single session because the physiological attributes being measured are subject to long term variability over and above the short term variability which would appear at a single session. Let  $x$  be the measurement vector and let  $H_i(x)$  be the distribution of  $x$  measured for the subjects at the  $i$ th data collection session. The long term distribution,  $F(x)$ , might be found by averaging the single session distribution,

$$F(x) = \frac{1}{N} \sum_{i=1}^N H_i(x) \quad (A47)$$

## DATA CAT STATISTICS

The distribution  $F$  is normalized if each  $H_i$  is normalized. Figure A5 shows how a set of different  $H_i$  can build an  $F$ . The vector  $x$  is shown as a scalar for ease of presentation.

We now postulate that the session-to-session variability is due to some hidden parameter  $y$ . For example, suppose variability in fingerprint measurements is caused by variability in skin moisture. Then  $y$  would measure moisture content. The postulate implies that for each  $y$  value ( $y$  is a vector), there is a unique  $H(x,y)$ . As different collection sessions are conducted,  $y$  will vary in time according to an unmeasured law and will result in different  $H$  distributions. That is, if  $y(t_i)$  is the value of  $y$  at the time of the  $i$ th observation,  $t_i$ , then

$$H(x,y(t_i)) = H_i(x) \quad (A48)$$

Let  $G(y)$  be the temporal density function of  $y$ . That is  $G(y)dy$  is the probability that a random sample of the hidden parameter will produce a value between  $y$  and  $y + dy$ . Then

$$F(x) = \int_{-\infty}^{+\infty} G(y) H(x,y) dy \quad (A49)$$

A hidden variable  $y$  may always be postulated. One may take  $y$  as time itself, for example. However, the existence of a distribution  $G(y)$  which can be normalized is an assumption which we make. This assumption is equivalent to stating that the long term variability in the parameter  $x$  is bounded. Since time is bounded in the access control situation of interest to us here, the function  $G(y)$  must always exist. The trivial case is when no hidden parameter other than

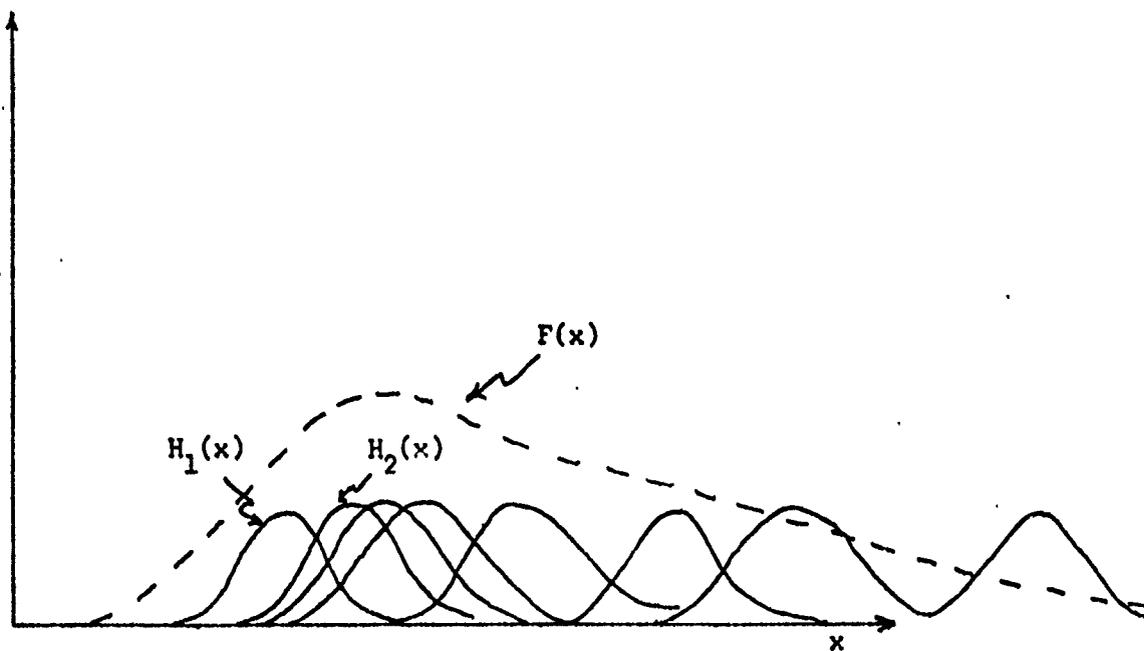


Figure A5 A set of  $H_i(x)$  observed at different sessions build a total distribution  $F(x)$  (not shown at same vertical scale)

## DATA CAT STATISTICS

time exists and  $G$  may be taken as the reciprocal of the time over which the data collection will occur,  $t_N - t_1$ . The interesting case is when  $y$  returns to the same value, making  $G$  a non-constant function of  $y$ .

In Section 1.0 we considered the problem of determining the number of subjects the data base should include. We used the notion of an undisclosed subgroup and, as an extreme example, considered that for one subgroup the Type I error  $p$  was the highest possible value, 1.0. We then argued that the Type I error quoted for a verification system should be the average over the population. A subgroup with a value of  $p = 1.0$  could comprise a small fraction of the population (namely 1%) and still permit the access device to meet specifications if the rest of the population had  $p = 0$ . If a data base were to contain few subjects, the probability of measuring the correct value of  $p$  for the population would be small since the sample would frequently contain too few or too many members of the poorly performing subgroup.

The results of Section 1.0 can be used to determine the number of data collection sessions required. In predicting the time averaged performance of an access control system, the worst case would arise when the hidden parameter  $y$  took on only two discrete values. When  $y = y_1$ , the subject has  $p = p_1 = 0.0$ . This case occurs 99% of the time, so  $G(y_1) = .99$ . However, when  $y = y_2$ , the system performance degenerates to a value of  $p = p_2 = 1.0$ , with  $G(y_2) = .01$ . The time

## DATA CAT STATISTICS

averaged Type I performance is .01% and still meets specifications. The hidden parameter  $y$  which governs the temporal variability of  $x$  is now analogous to the hidden subgroup which governs variability over subjects. We can thus state that 396 different data collection sessions are required to assure that any hidden parameter  $y$  does not possess statistics which will make the predicted Type I error erroneous. These sessions must, of course, be collected at times separated by an interval such that  $y$  will have a high probability of changing.

The number of sessions is appallingly large. However, by making some reasonable assumptions, the number can be reduced. If we assume the temporal variability in measured attribute,  $x$ , is the same for all subjects (only one  $G(y)$ ) but is uncorrelated in time between different subjects, then the result over many sessions can be inferred from the results over many subjects. For example, suppose we have 400 subjects who are enrolled at one session and tested at a later session with a time interval long compared to the time for variation in  $x$ . If there were a hidden parameter with the .01 probability of occurrence and  $p = 1.0$ , then the most probable occurrence is for 1% of the subjects to be rejected. As we show in Table A3, the 400 subjects permit determining of  $p$  of .01 with almost 90% confidence. Thus, if one satisfies the requirement on number of subjects, he will also satisfy the requirement on sessions if two sessions (counting enrollment) are used.

APPENDIX A  
REFERENCES

1. Cramer, Harold, "The Elements of Probability Theory", Wiley, New York, 1955.
2. Abramowitz, M. and Stegun, A., Handbook of Mathematical Functions, Dover, New York, 1968.
3. Sherwood, Bruce, A., "The Computer Speaks", Spectrum, p. 18, August 1979.
4. Oshika, B.T., "FACP Speech Recognition/Transmission System", RADC-TR-78-193, 1978.
5. Eleccion, Marce, "Automatic Fingerprint Identification", Spectrum, p. 36, September 1973.
6. Doddington, G.R. and Hydrick, B.M., "Speaker Verification", RADC-TR-75-274, 1975.
7. Benson, Peter, "Test Results, Advanced Development Models of BISS Identity Verification Equipment, Volume IV, Automatic Fingerprint Verification", Mitre Corporation, MTR-3442, September 1977.
8. Fejfar, Adolph, *ibid*, Volume III, "Automatic Handwriting Verification."

## APPENDIX B

### PAR SPEECH PROCESSING (PSP) SYSTEM

The PAR Speech Processing System is a flexible and easily expandable system within which a variety of speech processing tasks have been implemented. Data is stored in files in established formats, processing is carried out by independent tasks operating on these files, each implementing a single function.

There are four basic types of files: waveform files, containing digital speech data; encoded data files, containing linear prediction encoded speech data; phoneme library files, containing the phonemes used in construction; and covariance files, containing covariance matrices for phonemes used in phoneme recognition. Figure B1 shows the different file types and lists the programs and functions as they are related to the files. The following is a short description of each program.

Record: This task digitizes an analog speech signal using the LPA11-K(\*). The sampling rate is 12.8 kHz and has 12 bit (+/- 2048)

- - - - -

(\*) DEC PDP 11 series laboratory peripheral

## WAVEFORMS

Record:	Digitize speech signal
Playback:	Convert digital waveform to analog signal
Encode:	Encode speech into cross-sectional areas
Edit:	Extract portions of a waveform file
Display:	Display raw or processed waveform
Dump:	List out data values
Scale:	Scale waveform to 12 bits

## ENCODED DATA

Decode:	Decode cross-sectional areas into digital waveform
Change:	Modify voicing, pitch parameters
Display:	Display time history of cross-sectional areas
Dump:	List out data values
Construct:	Construct an encoded utterance

## PHONEMES

Enter:	Enter a phoneme into a library
Delete:	Delete a phoneme from a library
Dump:	List the phonemes in a library
Average:	Average many frames and enter into a library.

## COVARIANCE MATRICES

Covariance:	Calculate a covariance matrix for a file
Invert:	Invert the covariance matrices in a file
Classify:	Preliminary classification
Dump:	List entries in a covariance file
Delete:	Delete entries in a covariance file

Figure B1

## PAR SPEECH PROCESSING (PSP) SYSTEM

accuracy. Data is stored in waveform files, as 2 byte unformatted integers, 512 bytes/block. The start and stop of digitization is under operator control and the duration is limited only by the largest contiguous space on disk.

**Playback:** This task plays a digital waveform back out through the LPA11-K at 12.8 kHz. Start of D/A conversion is under operator control. The file can be auditioned repeatedly or a new file can be auditioned.

**Encode:** This task encodes a digital speech signal into linear prediction coefficients. The output file contains a frame label, if known, frame voicing, pitch period (if voiced), gain factor, fifteen linear prediction coefficients, fifteen reflection coefficients, and fifteen cross-sectional areas. These are 4 bytes, unformatted.

The encoding employs the auto-correlation method, as explained in Section 3.3.3. The computation is carried out using Robinson's recursion [1]. The reflection coefficients which are intermediate results of this calculation are used to calculate the cross-sectional areas using

$$\frac{A_m}{A_{m-1}} = \frac{1+k_m}{1-k_m}, \quad m = 1, 2, 3, \dots, M$$
$$A_M = 1$$

where  $A$  are the cross-sectional areas and  $k$  are the reflection coefficients.

## PAR SPEECH PROCESSING (PSP) SYSTEM

Voicing is detected using a cyclic auto-correlation, which is calculated by taking the inverse fourrier transform of the power spectrum. This function,  $r_c(n)$ , is searched for its maximum between  $n=2$  and  $n=256$ . If  $r_c(n_p)/r_c(1) \geq t_n$ , where  $t_n = 0.35$  and  $n_p$  is the location of the peak value, then the frame is called voiced, with a pitch period of  $P = n_p/f_s$ ,  $f_s$  is the sampling frequency, otherwise, the frame is called unvoiced with  $P = 0$ .

**Edit:** This task allows a section of a waveform file to be extracted and placed in another file. This is useful in eliminating the long silences before and after utterances, and in selecting short portions of long utterances for processing.

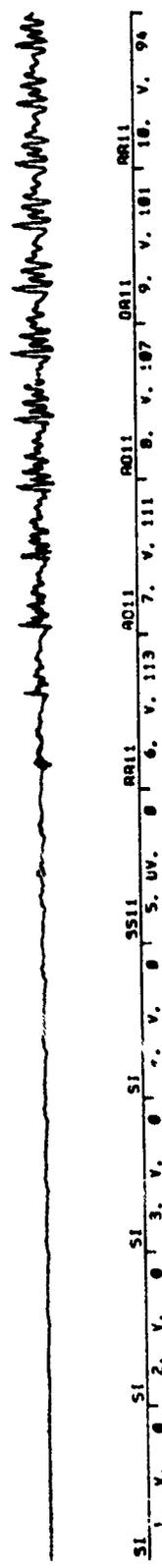
**Display:** Waveform files can be displayed on a Tektronix 4014 storage tube display terminal in two formats. The raw data can be displayed, with only the frame boundaries and frame numbers marked. If the file has a corresponding encoded data file, then the frame label (if known), voicing, pitch period, and frame number are displayed. Both displays are 10 frames/line, 4 lines/page (see Figure B2 and B3).

**Dump:** This task simply prints out the actual data values contained in a waveform file (Figure B4).

**Scale:** This task scales data from greater than 12 bits to 12 bits. It does not scale data up from less than 12 bits.

FILES: 000011.MOV

000011.ENC



SI 1. V. 0 2. V. 0 3. V. 0 4. V. 0 5. UV. 0 6. V. 113 7. V. 111 8. V. 107 9. V. 101 10. V. 94



ROI1 11. V. 90 12. V. 85 13. V. 83 14. V. 84 15. V. 86 16. UV. 0 17. UV. 0 18. UV. 0 19. UV. 0 20. UV. 0



TMI1 21. UV. 0 22. UV. 0 23. V. 106 24. V. 110 25. V. 110 26. V. 110 27. V. 108 28. V. 100 29. V. 106 30. V. 105



ROI1 31. V. 104 32. V. 107 33. V. 100 34. V. 107 35. V. 106 36. V. 105 37. V. 106 38. V. 106 39. V. 106 40. V. 107

Figure 12.

FILES: M00N1.001V

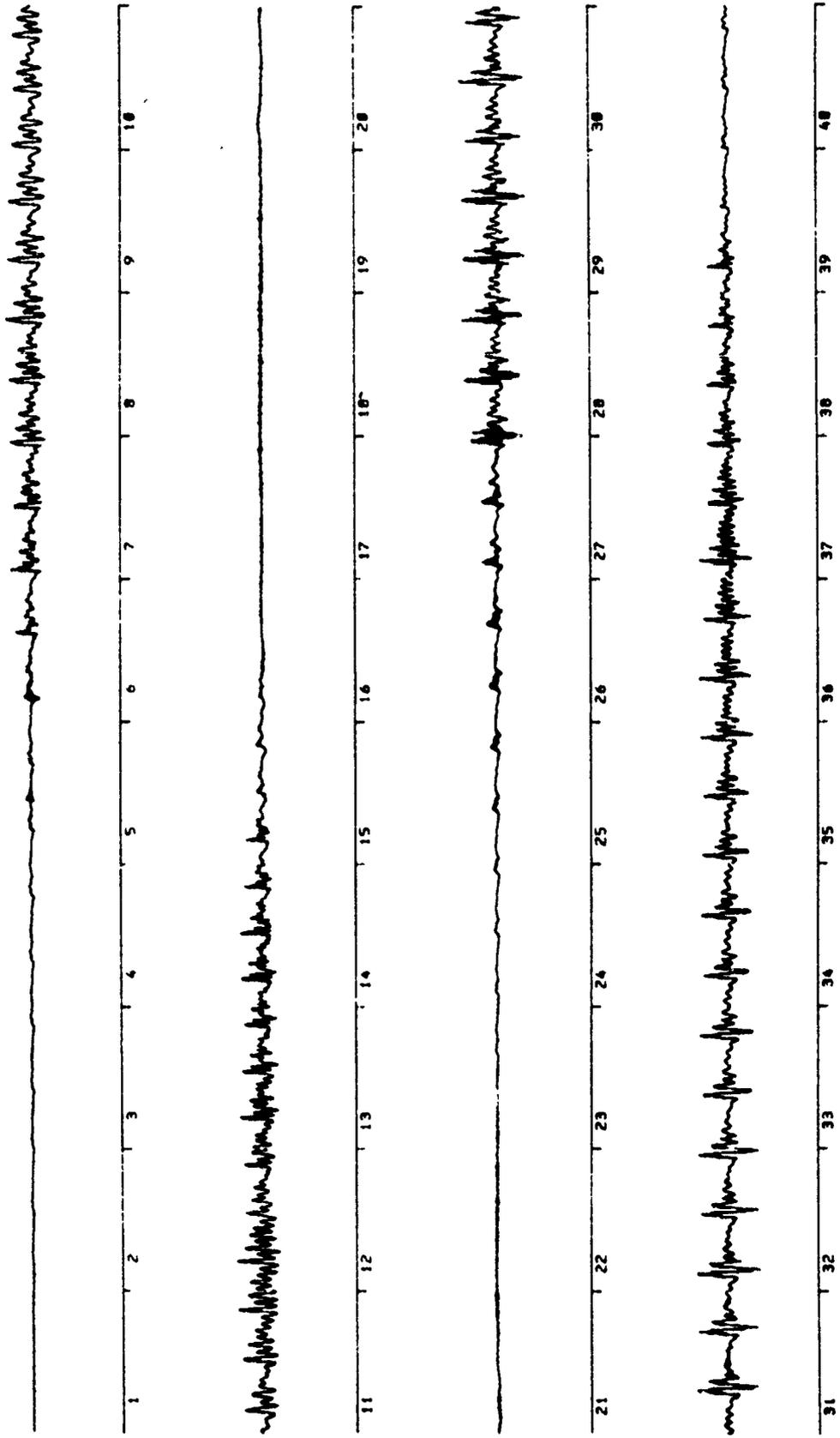


Figure B3.

DUMPING FILE: PRML1.UAU  
 HEADING:  
 SPQR:R NELSON, TI MATRIX, 14 AUG 80

RECORD:	28	-583	-116	226	-393	-203	487	-47	-184	284	164	-162	44	43	-119	-71	
	-64	-92	1	174	111	82	169	88	122	8	-124	-106	-89	-280	-270	-143	
	88	-161	-16	51	59	174	168	105	126	137	48	4	-32	-25	21	-38	
	-138	94	43	11	26	8	-32	-63	-106	-106	-77	-42	-12	-9	77	105	
	65	76	76	38	-9	-62	-71	-66	-97	-89	-60	-38	-158	-7	-31	-23	
	-18	-86	-120	-113	-131	-152	-162	-144	-120	-150	-122	-130	-6	-101	318	37	
	-116	425	298	-84	160	611	142	-366	94	-47	-503	-397	118	-520	179	57	
	3	-367	3	325	-28	-76	-266	-296	-52	-146	-187	-80	140	49	124	219	
	136	17	200	-152	-28	-124	12	-135	-97	6	-12	-10	75	52	78	72	
	172	112	-49	57	-49	-118	12	11	111	29	90	86	83	59	21	-8	
	20	-17	-84	-64	-63	-48	-57	-46	7	-18	-30	-35	-50	-79	-63	-99	
	-37	-72	-97	-101	-74	-100	-140	-176	-163	-238	63	230	-145	82	698	393	
	-101	-99	522	-171	-112	61	-300	-583	-191	-227	-543	-98	262	-150	-62	421	
	-72	454	114	215	-228	-71	12	-15	-191	101	56	-39	122	-175	26	-84	
	271	-49	114	-342	-1	-66	-73	-90	84	190	171	188	214	112	90	80	
	-184	-106	-153		-231												

TT1 -- STOP -- END OF DUMP

>

Figure B4.

## PAR SPEECH PROCESSING (PSP) SYSTEM

**Decode:** This task is complementary to the Encode task in that it creates a digital waveform from an encoded data file. The linear prediction coefficients are used to design a digital filter whose excitation is a pulse train for voiced speech or Gaussian distributed random noise for unvoiced speech.

**Change:** This task allows the user to modify the voicing decision and/or pitch period for any frames in an encoded data file.

**Display:** This task displays the time history of any one of the fifteen cross-sectional areas as a bar graph. There are ten frames per line, 4 lines per page, and the display is labeled with the frame label (if known), the voicing, pitch period, and frame number (Figure B5).

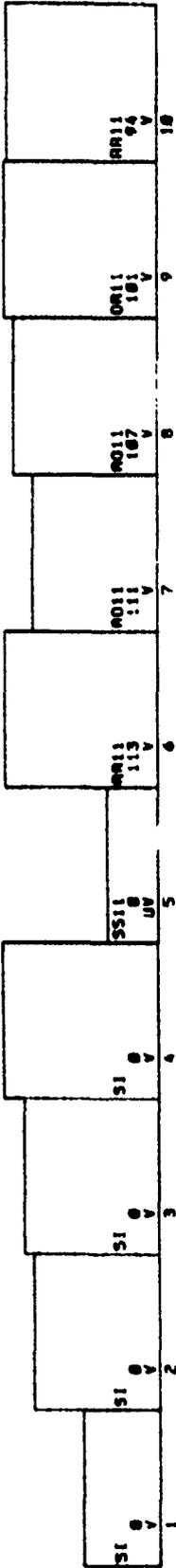
**Dump:** This task lists out the data contained in an encoded data file, frame by frame (Figure B6).

**Construct:** This task constructs an encoded data file according to a string of phonemes specified by the user. The data values used to construct the string are gotten from the appropriate phoneme name, a relative factor for the pitch and gain, and the duration. Control of the pitch and gain and duration gives the user control of the prosody of the utterance.

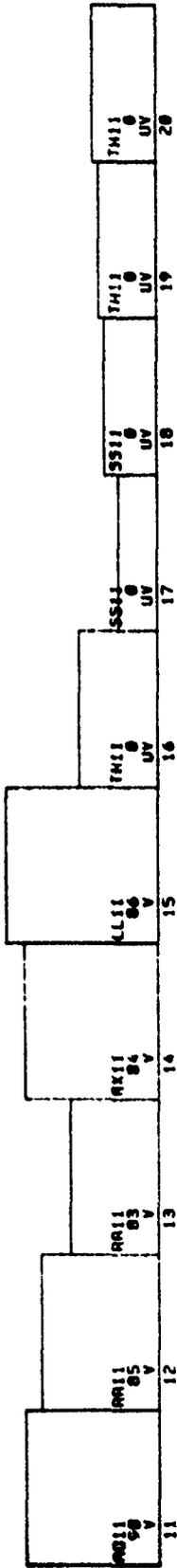
FILE: MMML1.ENC

TT1 --- STOP --- DISPLAY COMPLETED

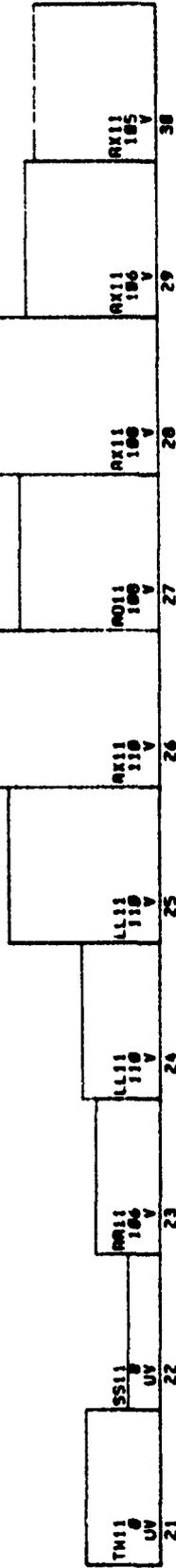
AC-14



AC-14



AC-14



AC-14

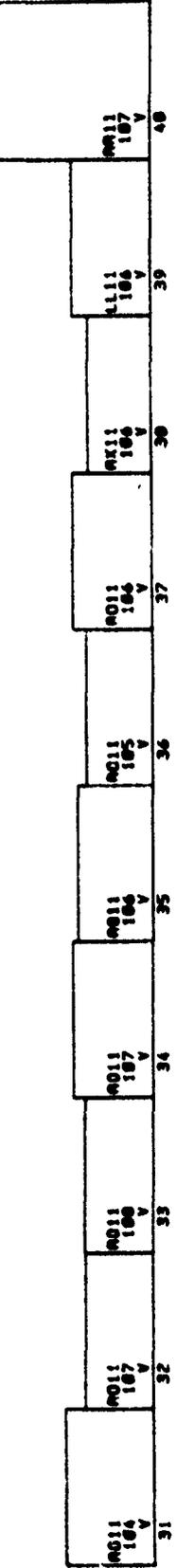


Figure B5.

DUMPING FILE: NRML1.ENC  
 HEADING: SPOR:M. NELSON, FIRST LINE OF TI MATRIX, EDITED FROM TIMRN.UAU, 16 AUG 80

RECORD: 28	U	108	511.024	0.56	1.63	0.33	-0.05	-1.10	-0.49	0.20	0.35	0.52	-0.07
AX11		-0.99	-1.95	0.66	0.64	0.16	0.26	-0.40	-0.17	-0.13	0.31	0.54	-0.07
		-0.42	-0.50	0.38	-0.63	0.16	0.26	1.29	3.02	4.26	5.57	2.92	0.88
		0.63	9.80	29.61	13.23	3.03	2.21						
		28.18	19.27	4.02	14.03								
RECORD: 29	U	106	609.744	0.64	1.05	0.54	-0.25	-0.71	-0.27	-0.02	0.44	0.27	0.05
AX11		-0.78	-0.88	0.15	0.55	0.19	-0.08	-0.21	-0.15	-0.08	0.42	0.27	0.05
		-0.68	-0.06	0.41	0.23	2.55	1.61	1.88	2.90	3.96	4.65	1.91	1.11
		5.27	27.18	30.82	12.87	8.05							
		21.42	19.42		8.05								
RECORD: 30	U	105	535.734	0.66	1.09	0.13	-0.15	-0.85	-0.14	0.26	0.27	0.39	-0.13
AX11		-0.87	-0.63	0.22	0.66	0.20	-0.00	-0.42	-0.02	0.09	0.37	0.39	-0.13
		-0.67	0.15	0.36	0.34	2.96	1.97	1.98	4.88	4.65	3.87	1.77	0.78
		9.52	47.53	35.37	16.57	8.21							
		14.83	21.51		16.57								
RECORD: 31	U	104	494.080	0.80	0.91	0.08	-0.21	-0.77	0.03	0.16	0.34	0.16	-0.03
AX11		-0.82	-0.61	0.22	0.80	0.17	-0.02	-0.46	0.12	0.16	0.36	0.16	-0.03
		-0.62	0.04	0.48	0.39	2.47	1.75	1.83	4.93	3.83	2.78	1.30	0.94
		9.73	41.33	37.90	13.42	5.86							
		11.58	19.05		13.42								
RECORD: 32	U	107	430.353	0.81	0.56	-0.04	-0.39	-0.56	0.20	0.17	0.27	0.05	-0.04
AX11		-0.77	-0.48	0.45	0.56	0.04	-0.27	-0.28	0.27	0.22	0.29	0.04	-0.04
		-0.62	0.31	0.58	0.35	3.03	1.58	2.74	4.92	2.81	1.79	1.00	0.92
		10.53	45.03	23.59	6.36	1.72							
		7.63	17.42		6.36								

TT1 -- STOP -- END OF DUMP

>

Figure B6.

## PAR SPEECH PROCESSING (PSP) SYSTEM

The construction program reads the phonemes out of the library by pairs. Starting with the first and second, it first duplicates them for the duration specified in a buffer. Then transitions are calculated for the cross-sectional areas, gain and pitch. Then new values are calculated for the linear prediction and reflection coefficients are calculated as the first phoneme is written to the output encoded data file, frame by frame. The third phoneme is then read in, duplicated, and transitions between it and the second phoneme are calculated. The second phoneme is output, and the procedure repeats until the last phoneme is output. Such a constructed utterance can then be decoded and auditioned. This is the speech synthesis task. Figure B7 shows the utterance construction processing.

**Enter:** Phoneme values can be selected from an encoded data file and inserted into a library.

**Delete:** Phoneme entries in a library can be deleted.

**Dump:** The phoneme entries in a library are listed out by this task (Figure B8).

**Average:** Many frames of an encoded data file are averaged by this task and entered into a library as a phoneme. This is useful for phonemes that can be made as sustained sounds, such as vowels, nasals and fricatives. A waveform consisting of only one sustained phoneme

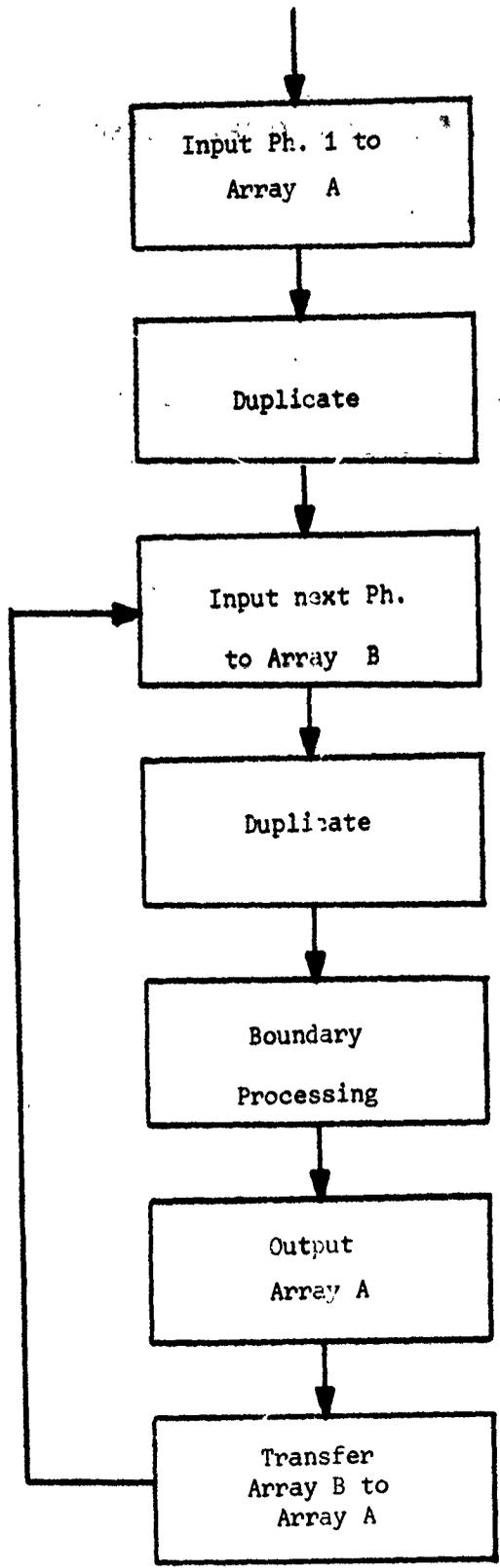


Figure B7.

HERE IS THE CATALOG OF ENTRIES FOR: TIMRN.LIB

SPKR:M.	NELSON, PHONEMES FOR TI MATRIX, CREATED 15 AUG 80			
NH11	TIMRN.ENC	34	T	120
OR11	TIMRN.ENC	4	T	0
RR11	TIMRN.ENC	44	T	84
TH11	TIMRN.ENC	48	F	0
SI	TIMRN.ENC	25	T	0
OR21	TIMRN.ENC	40	T	94
SI1	TIMRN.ENC	25	T	0
LL11	TIMRN.ENC	54	T	110
AO11	TIMRN.ENC	62	T	107
NA21	TIMRN.ENC	71	T	110
GG11	TIMRN.ENC	79	T	107
RR21	TIMRN.ENC	83	T	112
EY11	TIMRN.ENC	86	T	108
TT11	TIMRN.ENC	91	T	0
CC11	TIMRN.ENC	98	T	27
AE11	TIMRN.ENC	104	T	121
PP11	TIMRN.ENC	117	F	0
PP12	TIMRN.ENC	118	T	29
GG21	TIMRN.ENC	78	F	0
TT21	TIMRN.ENC	239	F	29
CC21	TIMRN.ENC	99	F	0
TT31	TIMRN.ENC	284	F	0
NH11	TIMRN.ENC	106	T	126
PP21	TIMRN.ENC	579	F	0
PP22	TIMRN.ENC	580	F	0
CC31	TIMRN2.ENC	707	F	0
NA21	TIMRN2.ENC	546	T	101
PP31	TIMRN2.ENC	955	F	0
PP32	TIMRN2.ENC	956	F	0
TT1	STOP			

-- END OF LIBRARY CATALOG

## PAR SPEECH PROCESSING (PSP) SYSTEM

is encoded, then averaged, then entered into the library.

**Covariance:** A library containing at least fifteen different occurrences of the same phoneme is used by this task to calculate a covariance matrix for that phoneme and enter it into a covariance matrix file, along with the mean value.

**Invert:** This task inverts the covariance matrices in a covariance matrix file and generates a file in the same format, but with the inverted matrices in place of the covariance matrices. The matrices are stored in upper triangular column form since they are symmetric.

**Dump:** This task lists out the entries of a covariance or inverted matrix file (Figure D9).

**Delete:** This task deletes entries from a covariance or inverted matrix file.

**Classify:** This task uses the inverted covariance matrix file to nominate phoneme names for each frame of an encoded data file, using a Mahalanobis weighted nearest mean vector logic [2]. The phoneme names are inserted into the label fields of encoded data file. The user can use these as a guide in making the phoneme selection for entry into the library.

LABEL: NM11 VOICE: T PITCH: 93 GAIN: 157.

COU1

0.2973258E-02	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.3388994E-02	0.4805510E-02	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
-0.7831138E-04	0.3552854E-03	0.426202E-03	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.8819106E-03	0.3063386E-02	0.1988695E-02	0.000000	0.1236189E-01	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.8452148E-03	0.1695091E-02	0.3684997E-03	0.000000	0.1907570E-02	0.1329919E-02
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.1753341E-02	0.4117981E-02	0.2081720E-02	0.000000	0.1281649E-01	0.2204403E-02
0.1691620E-01	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
-0.5785027E-03	0.9106328E-03	0.9263625E-03	0.000000	0.4040409E-02	0.2328324E-02
0.4218062E-02	0.6632220E-02	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.2122012E-02	0.5070317E-02	0.2135049E-02	0.000000	0.1428114E-01	0.2585674E-02
0.1753570E-01	0.4305517E-02	0.2174623E-01	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.2485627E-02	0.7892646E-02	0.3825274E-02	0.000000	0.242306E-01	0.5484061E-02
0.2914225E-01	0.1379955E-01	0.2789715E-01	0.000000	0.7069981E-01	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.3972020E-02	0.7559429E-02	0.1286069E-02	0.000000	0.2724444E-02	0.6648351E-02
0.5987816E-02	0.1014608E-01	0.825795E-02	0.000000	0.8953215E-02	0.5104535E-01
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.1548493E-02	0.5678375E-02	0.2654309E-02	0.000000	0.1429148E-01	0.3798638E-02
0.1859073E-01	0.7632367E-02	0.1683353E-01	0.000000	0.3703333E-01	0.1017534E-01
0.2525995E-03	0.000000	0.000000	0.000000	0.000000	0.000000
0.3658959E-02	0.9271985E-02	0.3540169E-02	0.000000	0.1988115E-01	0.5176335E-02
0.2657541E-01	0.9690366E-02	0.2726579E-01	0.000000	0.4898849E-01	0.2791612E-01
0.2668107E-01	0.7165705E-01	0.000000	0.000000	0.000000	0.000000
0.4750011E-02	0.8595867E-02	0.1457877E-02	0.000000	0.5878206E-02	0.5622589E-02
0.5462088E-02	0.6913363E-02	0.6780261E-02	0.000000	0.1496589E-01	0.2740236E-01
0.1881436E-01	0.7399154E-02	0.4489319E-01	0.000000	0.000000	0.000000
-0.2293720E-03	0.4616878E-03	0.1355209E-02	0.000000	0.5716407E-02	0.5372138E-03
0.6324631E-02	0.1355334E-02	0.5028241E-02	0.000000	0.9965511E-02	0.1461443E-02
0.6990341E-02	0.9428007E-02	0.1764359E-02	0.000000	0.7712809E-02	0.000000
0.2485649E-02	0.5151266E-02	0.1575539E-02	0.000000	0.8434587E-02	0.2013694E-02
0.1110210E-01	0.4750136E-02	0.112169E-01	0.000000	0.2072740E-01	0.1418036E-01
0.1036231E-01	0.2220264E-01	0.1457593E-01	0.000000	0.3373589E-02	0.1256640E-01

## PAR SPEECH PROCESSING (PSP) SYSTEM

Any of these tasks may be altered without affecting the file structure or other tasks and any new tasks may be added, using the same files, and/or creating any new files needed. This is the key to flexibility and extensibility. Figure B10 shows the general processing flow in this system.

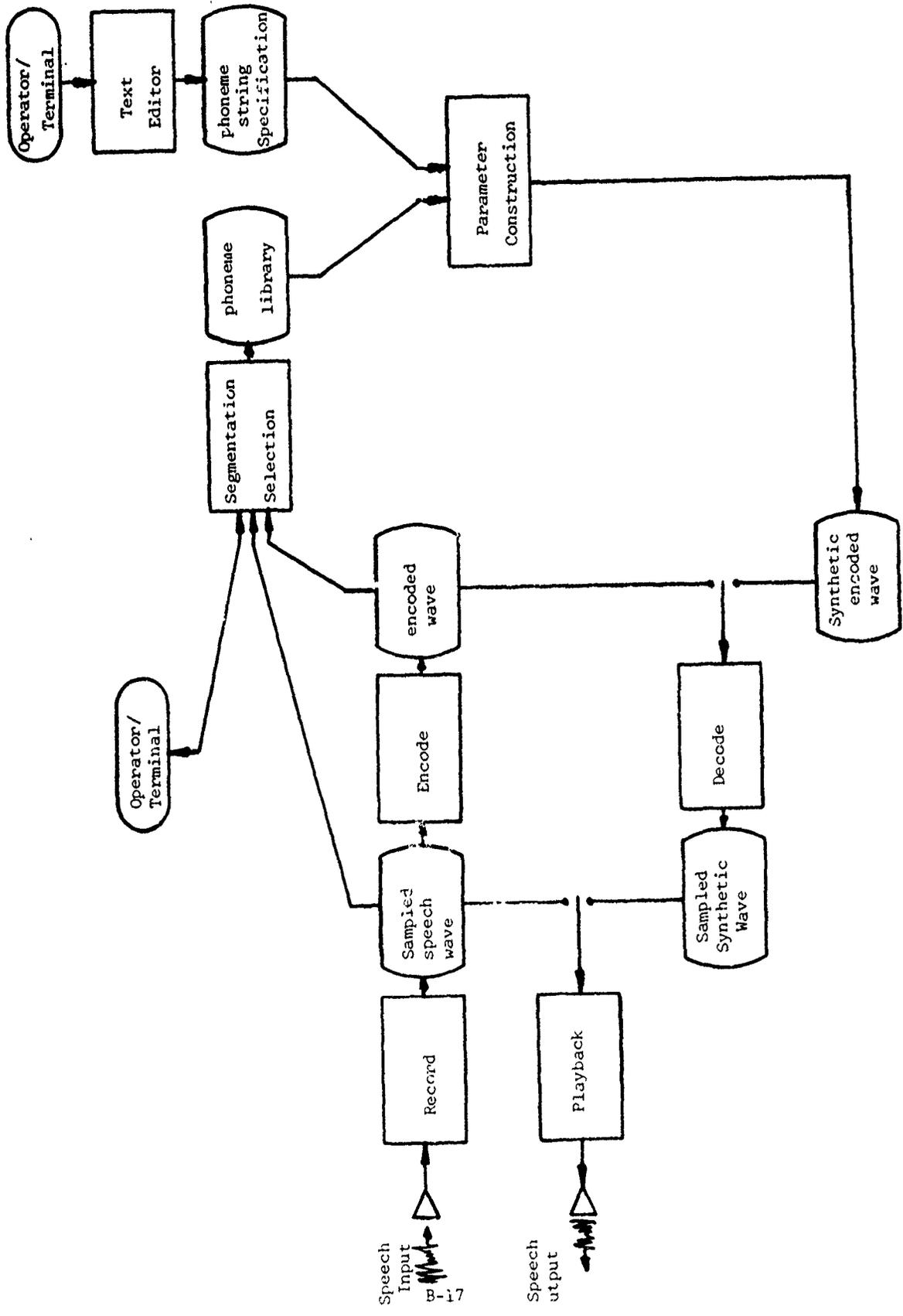
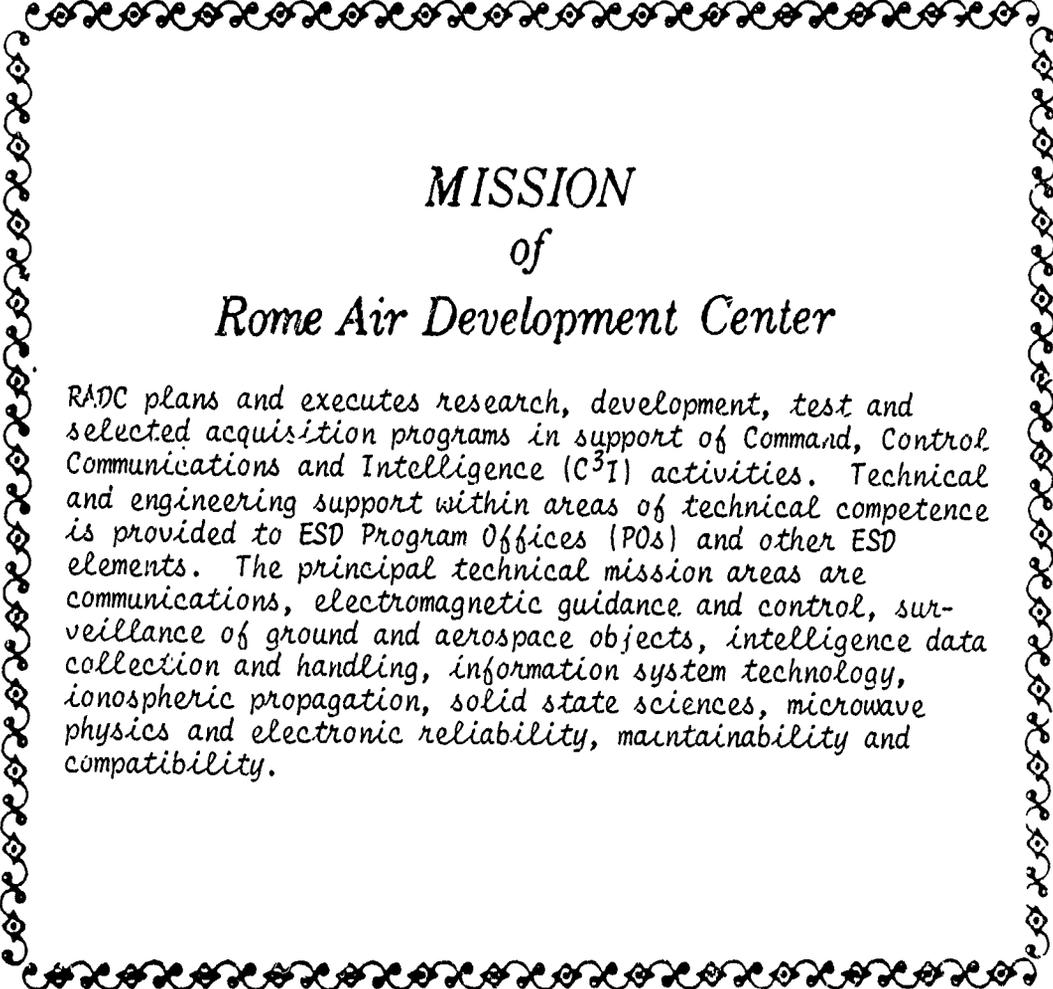


Figure B10.

APPENDIX B  
REFERENCES

- [1] J. D. Markel, A. H. Gray, Linear Prediction of Speech, Springer-Verlag, 1976.
- [2] R. O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley-Interscience, John Wiley & Sons, 1973.



*MISSION*  
*of*  
*Rome Air Development Center*

*RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control Communications and Intelligence (C<sup>3</sup>I) activities. Technical and engineering support within areas of technical competence is provided to ESD Program Offices (POs) and other ESD elements. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.*