AD-A050 841 TEXAS INSTRUMENTS INC DALLAS F/G 5/7

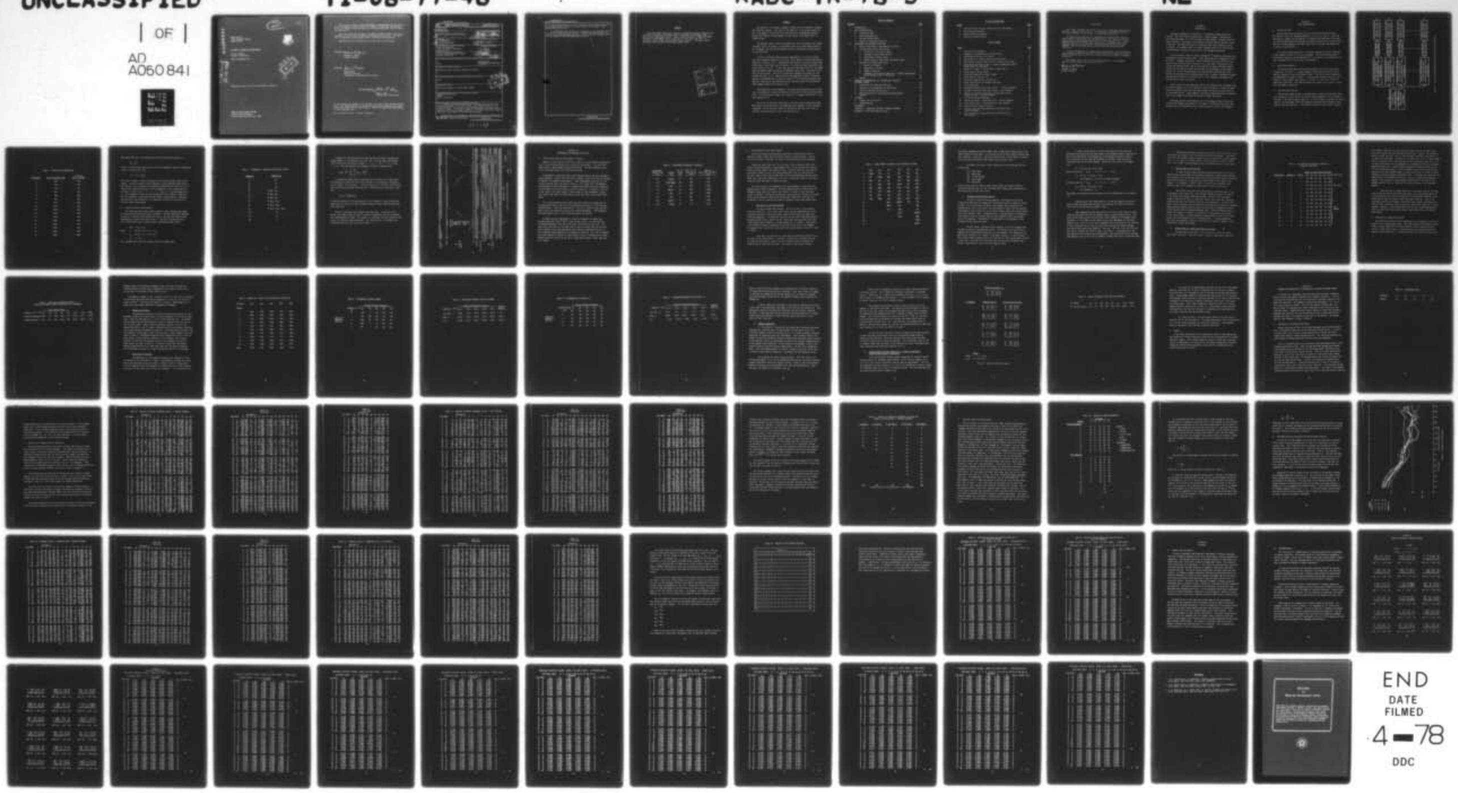
AUTOMATIC LANGUAGE DISCRIMINATION.(U)

JAN 78 R G LEONARD, G R DODDINGTON F30602-76-C-0168

UNCLASSIFIED TI-08-77-46 RADC-TR-78-5 NL

| OF |

AD
A060841



END
DATE
FILMED
4-78
DDC

RADC-TR-78-5
Final Technical Report
January 1978

AUTOMATIC LANGUAGE DISCRIMINATION

R. Gary Leonard
George R. Doddington

TEXAS INSTRUMENTS INC.

Approved for public release; distribution unlimited.

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

This report contains a large percentage of machine-produced copy which is not of the highest printing quality but because of economical consideration, it was determined in the best interest of the government that they be used in this publication.

This report has been reviewed by the RADC Information Office (OI) 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-78-5 has been reviewed and is approved for publication.

APPROVED: Richard S. Vonusa

RICHARD S. VONUSA
Project Engineer

APPROVED: Howard Davis

HOWARD DAVIS
Technical Director
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)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: RADC-TR-78-5	2. GOVT ACCESSION NO.
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): AUTOMATIC LANGUAGE DISCRIMINATION	5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report, Mar [redacted] 76 - Aug [redacted] 77
	6. PERFORMING ORG. REPORT NUMBER: TI-08-77-46
7. AUTHOR(s): R. Gary Leonard, George R. Doddington	8. CONTRACT OR GRANT NUMBER(s): F30602-76-C-0168
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Texas Instruments Inc., 13500 North Central Expressway, Dallas TX 75222	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: P.E. 31011GF, J.O. 70550723
11. CONTROLLING OFFICE NAME AND ADDRESS: Rome Air Development Center (IRAA), Griffiss AFB NY 13441	12. REPORT DATE: Jan [redacted] 78
	13. NUMBER OF PAGES: 71
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Same	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: Richard Vonusa (IRAA)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Language Classification
Speech
Pattern Recognition

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A language discrimination algorithm was developed that utilizes single steady-state phoneme-like reference sounds. The algorithm makes use of the relative frequency of occurrence of these reference sounds and nearest-mean decision strategy. The reference sounds were determined interactively and have been shown to have good language specificity. The number of reference sounds required is small enough to allow real-time operation.

Standardization of each speaker's long-term average spectrum removed much

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

347650

of the data variation that was due to variations in data recording conditions. This standardization provided a 41 percent decrease in the number of test speaker misclassifications.

In a performance test involving 50 test speakers of five languages, 80 percent correct classification was achieved. Excellent discrimination among L1, L3, and L5 was attained, while additional references with more language specificity are needed for discriminating between L2 and L4.

PREFACE

This Final Report describes research on automatic language classification by Texas Instruments Incorporated, Central Research Laboratories, 13500 North Central Expressway, Dallas, Texas, under Contract No. F30602-76-C-0168 for Rome Air Development Center, Griffiss Air Force Base, New York. Mr. Richard S. Vonusa (IRAP) was the RADC Project Engineer. The report covers work performed from March 1976 through August 1977.

SUMMARY

The problem is to develop a computer simulation of an automatic language discrimination machine. The input to the simulator is an analog speech signal. The output is a decision as to the particular language being spoken. Neither the identity of the speaker nor the nature of the spoken text is assumed known.

The approach taken to solve the problem was to first determine key sounds (phonemes, words, or phrases, as examples) which are to some degree specific to a language, and then use estimated occurrence frequencies of these sounds to make decisions.

A study of consonant-vowel-consonant hyperphones in English indicated that such 3-phoneme sequences do not occur often enough to provide operational capability (Section III.A). Consideration was given to the use of generalized sound types, rather than specific sounds. Six types of sound transitions were studied (Section III.B). It was found that the continuous nature of the speech data precluded reliable detection of component transitions, using 50 ms reference representations. The use of steady-state phoneme-like sounds was considered (Section III.C). Departing from the completely automatic reference selection techniques used in previous studies,[1,2] interactive procedures were applied, resulting in the extraction of 36 reference sounds which possessed some language specificity (Section IV.B).

Standardization of each speaker's long-term average spectrum was carried out in an attempt to reduce the effects of variations in data recording conditions. This procedure allowed a 41% reduction in language classification error rate (Section IV.C).

The use of 13 reference sounds plus a voicing interval measure provided 80% correct classification of 50 independent test speakers of five different languages. Due to the small number of references required, this level of accuracy can be achieved in real time (Section IV.D).

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

EVALUATION

This report discusses the latest in a series of experiments dealing with automatic language classification. The techniques and methods developed in this study resulted in significantly increased accuracy results.

An algorithm was developed for language discrimination that utilized single steady-state phoneme-like reference sounds. The required reference sounds were of a small enough number to allow real-time operation. A set of reference data which was developed interactively, indicated a significant language specificity. Accuracy was improved through standardizing each speaker's long-term average spectrum to minimize recording conditions variations.

In addition, measures of language documentation other than the phoneme-like reference sounds were utilized. Future consideration should be given to prosodic measures such as voice to unvoiced intervals as a means of language recognition.

Performance tests involving fifty test speakers of five languages resulted in 80 percent correct classification.

Richard S. Vonusa

RICHARD S. VONUSA
Project Engineer

SECTION I

INTRODUCTION

Automatic language classification is a challenging speech processing problem. The three classic problem areas, unknown speaker, unknown text, and connected speech, are intrinsic parts of the automatic language classification problem. Fortunately, many seconds of speech data may be used to form a decision. The ideal strategy is to determine key sound sequences (words or phrases, for example) which are highly language specific and then perform limited sequence recognition and classify the language accordingly.

In two previous studies at Texas Instruments[1,2] automatic techniques were developed for selecting key sounds. Those techniques were independent of the particular languages considered. Results obtained during the course of the present study and described in Section III of this report indicated that an interactive approach might lead to the selection of useful reference sounds. A small set of reference sounds was interactively developed, as described in Section IV. Results of processing the training data indicated that these sounds did have significant language specificity. However, subsequent processing and classification of the test data in a five-language, 50-speaker test yielded 66% correctly classified.

Close scrutiny of the processing results and comparison of the speech data of several speakers led to the conjecture that differences in recording conditions and equipment might be significant enough to cause the large variation in processing results. Each speaker's long-term average spectrum was standardized in an attempt to minimize the results of these variations.

Following spectral normalization, the five-language classification test involving 50 speakers was repeated. The number of classification errors was reduced by 41%, allowing 80% correct classification of the 50 test speakers.

SECTION II
DATA REPRESENTATION

A. Analog Data Base

Analog speech data recordings for this study were provided by RADC. The data are from five languages denoted L1, L2, L3, L4, and L5. Data from 100 distinct speakers were processed: 50 speakers provided data for estimating decision parameters and generating reference sound files (training data), and data from 50 speakers (test data) were used to estimate the probability of correct classification.

The training data consisted of five-minute segments of speech from each of ten speakers of each of the five languages. The test data consisted of five-minute segments from ten speakers of L1, L3, and L5; six speakers of L2; and 14 speakers of L4, for a total of 50 speakers.

B. Analog Preprocessing

The analog speech data base was preprocessed using the hardware shown functionally in Figure 1. Sixteen bandpass filters were used to provide a spectral analysis of the input speech signal. The filter center-frequencies and bandwidths are shown in Table 1. Following the low-pass filtering, the signals in each channel were sampled, digitized to 11 bits, and stored for additional processing. One hundred samples per second were retained to represent the speech information.

C. Raw Data Normalization

Some speaker normalization is accomplished by regressing each data vector upon regression vectors chosen to minimize between-speaker to within-speaker variance in the time-frequency spectrum. In the following, reference to specific time will be suppressed since the preprocessing operations performed at each sampling time are identical. Let D denote a vector of data values stored at

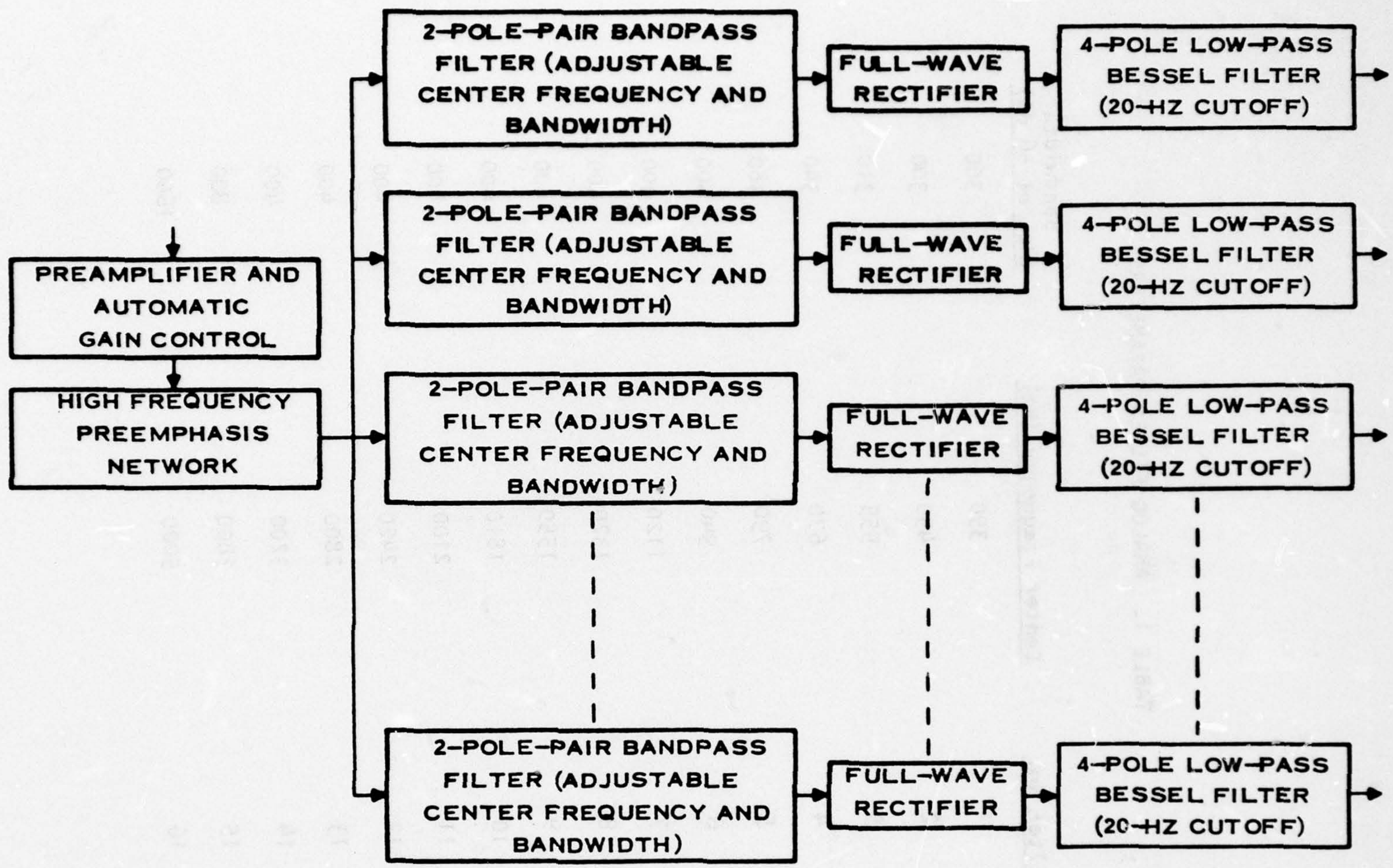


Figure 1 Functional Block Diagram, Analog Spectral Preprocessor

TABLE 1. ANALOG FILTER PARAMETERS

Filter No.	Center Frequency (Hz)	Bandwidth (Hz, at -6 dB)
1	350	300
2	450	300
3	555	310
4	670	340
5	790	380
6	940	400
7	1120	400
8	1320	400
9	1550	400
10	1810	400
11	2100	400
12	2420	400
13	2800	400
14	3200	400
15	3800	800
16	5000	1600

some specified time. The expression for the normalized data vector is

$$G_N = \frac{1}{\sigma} H \quad ,$$

where H is the original data vector D with the components along the regression vectors subtracted out, and

$$11\sigma = H' \cdot H = \sum h_i^2 \quad .$$

The value of sigma is used as a measure of the overall energy level of the speech. Letting g_i denote the ith component of the normalized data vector G and letting c_1 and c_2 denote the regression coefficients for the first- and second-order regression vectors, the ith component f_i of a reduced vector F is shown in Table 2 for i = 1, 2, ..., 12. Thus, F is a twelve-element vector, nine elements of which are from normalized filter outputs, along with the two regression coefficients and the overall energy measure. Following normalization, each element of F is quantized to three bits and stored on disk for rapid random access.

D. Reference Sound Representation

The reference sounds used in this study are short, phoneme-like sounds or transitions from one such sound to another. Let $\underline{S}(t)$ denote the representation of a reference sound centered at time t, and let F(t) denote a preprocessed data vector occurring at time t. Then $\underline{S}(t)$ is defined to be a 12 x 3 derived data matrix consisting of three derived data vectors:

$$\underline{S}(t) = [S_1 \; S_2 \; S_3],$$

where

$$S_1 = \tfrac{1}{2}\,[F(t - 2) + F(t - 1)],$$

$$S_2 = \tfrac{1}{2}\,[F(t + 1) + F(t + 2)],$$

$$S_3 = S_1 - S_2.$$

This representation spans five samples (50 ms) of speech data.

TABLE 2. COMPONENTS OF REDUCED-SPECTRUM DATA VECTOR

Component	Composition
f_1	g_1
f_2	g_2
f_3	g_3
f_4	$\frac{1}{2}(g_4 + g_5)$
f_5	$\frac{1}{2}(g_6 + g_7)$
f_6	$\frac{1}{2}(g_8 + g_9)$
f_7	$\frac{1}{2}(g_{10} + g_{11})$
f_8	$\frac{1}{2}(g_{12} + g_{13})$
f_9	$\frac{1}{3}(g_{14} + g_{15} + g_{16})$
f_{10}	c_1
f_{11}	c_2
f_{12}	σ

A measure of the dissimilarity of two matrices $\underline{S}^1$ and $\underline{S}^2$, representing sound segments, is the squared error $e(\underline{S}^1, \underline{S}^2)$. Let s_{ij}^k denote the element of $\underline{S}^k$ in the i-th row and j-th column, i = 1, 2, ..., 12; j = 1, 2, 3; and k = 1, 2. Then the squared error between $\underline{S}^1$ and $\underline{S}^2$ is defined to be

$$e(\underline{S}^1, \underline{S}^2) = \sum_{j=1}^{3} \sum_{i=1}^{12} (s_{ij}^1 - s_{ij}^2)^2 \quad .$$

In processing speech data to determine the recurrence of a fixed reference sound $\underline{S}$, the data are said to be "scanned" by $\underline{S}$, producing a "scanning error," $E(\underline{S},t)$, for each time t. Let $\underline{D}(t)$ denote a 3 x 12 matrix derived from the input data of the same format as for a reference sound, centered at time t. Then

$$E(\underline{S},t) \overset{D}{=} e\left(\underline{S},\underline{D}(t)\right) \quad .$$

A relative minimum in the scanning error for $\underline{S}$ indicates a time at which the speech data are similar to the reference sound $\underline{S}$ and hence flags the possible recurrence of $\underline{S}$.

Figure 2 illustrates the concepts described above. A portion of the phrase, "has taken steps to assure," has been preprocessed and the 12-element data representation printed. Also shown is a reference sibilant chosen at time t_0 and its resultant scanning pattern. The scanning error is plotted, and a scanning valley is seen to occur at time t_1.

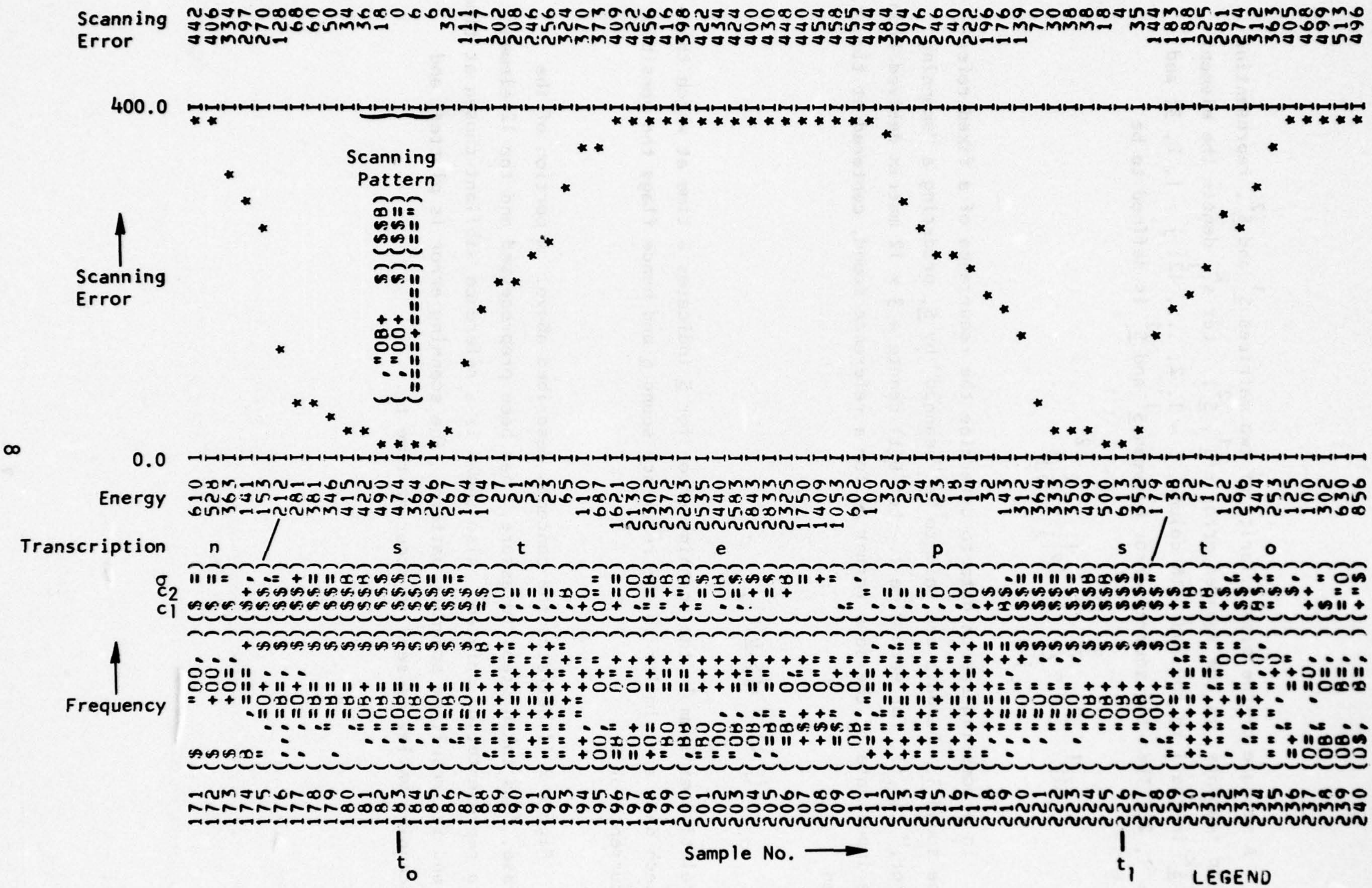


Figure 2 Data Processing Example

LEGEND

SYMBOL	,	"	+	=	0	8	$
LEVEL 0	1	2	3	4	5	6	7

SECTION III

BACKGROUND FOR REFERENCE SELECTION

A. Rate of Occurrence of Hyperphones in English

Results in this section arise from a study of a written English data base consisting of memos generated in an industrial environment. Of the 64,188 words written, approximately 6,000 were distinct. The most frequently occurring 1,024 words account for 36,772 written words.

A hyperphone is defined to be a sound sequence having the form consonant-vowel-consonant. Each of the 1,024 most frequently occurring words was analyzed to identify its component hyperphones. Those hyperphones which occur in ten or more distinct words are shown in Table 3. This table lists the phonemic representation of each hyperphone, an example word containing that hyperphone, and the number of distinct words in which it occurred. In addition, item (4) in the table is the total number of occurrences of each hyperphone, taking into account the number of occurrences of each word that contains that hyperphone.

If it were assumed that this data base were read at an average rate of 100 words per minute, then the first hyperphone would occur, on the average, 472(100)/36772 = 1.28 times each minute. Item (5) of Table 3 shows the estimated rate of occurrence for each of the listed hyperphones. The hyperphone with the highest rate is "for" (2.28 occurrences per minute).

In detecting such hyperphones it would be reasonable to expect the average false alarm rate to be at least one per minute. For the typically unstressed hyperphone "for" the false alarm rate would be higher yet. The false alarm rate being of the same order as the occurrence rate for even the most frequently occurring hyperphones would severely compromise the use of such 3-phoneme sound sequences as language discriminants. This result suggests that study should be limited to single sounds or pairs of phoneme-like sounds. Another possibility would be the use of broad sound classes, rather than specific sounds.

TABLE 3. HYPERPHONE OCCURRENCES IN ENGLISH

	Hyperphone Phonetic Rep. (1)	Example (2)	No. of Words (3)	Total No. of Occurrences (4)	Rate of Occurrence (No./Min) (5)
1	ʃIn	act<u>ion</u>	31	472	1.28
2	mIn	depart<u>ment</u>	18	470	1.28
3	ℓi-	on<u>ly</u>	18	282	0.767
4	-In	<u>en</u>sure	14	183	0.498
5	ri-	ve<u>ry</u>	14	296	0.804
6	for	<u>for</u>	11	837	2.28
7	kan	<u>con</u>cern	10	168	0.457
8	ðId	no<u>ted</u>	10	91	0.247

B. Classification Using Sound Classes

This section describes experiments and results concerning the use of classes of sounds rather than particular sounds or sound sequences. Before the sound classes are defined, two aspects of data processing are discussed.

Processing input data with a file of one or more reference sounds shall refer to the determination, at each sample time, of the scanning error corresponding to each sound in the reference file. Relative minima in the scanning error whose values are less than some prespecified value of scanning error, say EMAX, shall flag a time of recurrence of the corresponding reference sound. Hence, this recurrence acceptance criterion is relaxed when EMAX is relatively large and tight when EMAX is relatively small.

During processing, the smoothed overall energy measure is monitored and compared with a preset threshold. When the smoothed energy remains below this threshold longer than 0.25 second, occurrence of silence in the input data stream is assumed. During data processing, such periods of silence are ignored. To say that N seconds of data are processed shall mean that input data are processed as long as required to process N seconds of nonsilence speech data.

1. Definition of the Sound Classes

Six sound classes were studied. The first class is the vowel-to-fricative transition. Use of this sound class requires determination of the occurrence of such a transition without regard to the particular vowel and fricative components. To determine such occurrences, eight reference patterns representing particular vowel-fricative transitions were extracted from individual English words spoken by a single speaker (R. E. Kromer). Each pattern is an average of several utterances of the same transition by this speaker. The spoken words in which the transitions occurred are shown as the first item in Table 4.

Input data are processed to simultaneously determine scanning errors for each of the eight patterns. Typically, scanning error valleys occurred at about the same time for several of the reference patterns whenever the vowel-fricative transition occurred. This results from using a relaxed

TABLE 4. SOURCE WORDS FOR SOUND CLASS REFERENCE PATTERNS

	v-f	v-n	v-s	f-v	n-v	s-v
1	these	him	it	he	me	be
2	if	thing	get	she	men	bin
3	with	an	at	his	may	been
4	is	time	had	thing	not	take
5	this	some	not	said	must	they
6	as	one	such	say	now	tell
7	us	only	put	such	no	by
8	was	them	out	some	more	time
9			upon	should	any	but
10			about	for		put
11			up	who		to
12				first		people
13				see		upon
14						about
15						before

occurrence acceptance threshold (EMAX = 80), attempting to detect most of the vowel-fricative transitions. A collection of one or more scanning valleys that are clustered in time (i.e., that occur within 50 ms of each other) is considered to be one occurrence of the vowel-fricative sound class.

The other five sound classes studied are also transitions and are listed below:

(2) vowel-nasal
(3) vowel-stop
(4) fricative-vowel
(5) nasal-vowel
(6) stop-vowel

The L6 source words for each of these sound classes are shown in Table 4. Occurrence determination for the five additional classes was accomplished as described above for the first class.

2. Language Classification Results

For each of the 50 training speakers, 180 seconds of data were processed to determine the numbers of occurrences of each of these six sound classes. For classification purposes, each speaker was considered to be represented by the 6-tuple vector, the coordinates of which are the standardized numbers of occurrences of the six sound classes found during the processing of that speaker's data. Standardization is accomplished by transforming the vectors to have zero sample mean and unit sample variance as averaged over all 50 speakers. The modifier "standardized" will be assumed in the following unless otherwise noted.

The ten vectors representing the speakers of a given language were averaged to obtain a language mean vector. Classification of the training speakers was accomplished by choosing, for each speaker, the language of the mean vector to which it is closest in Euclidean distance. In what follows, this classification procedure shall be referred to as the nearest-mean classification rule. The result of classifying each of the training speakers was correct classification for 33 of the 50 speakers (66%).

A linear transformation V of the six-dimensional data space was determined which maximizes the mean-square interlanguage distance, keeping the sum of the mean-square interlanguage and intralanguage distances constant. This transformation then separates speakers of different languages while clustering those of the same language. Denote the data vector from speaker s by

$$F_s = (f_{s1}\ f_{s2}\ f_{s3}\ f_{s4}\ f_{s5}\ f_{s6}).$$

Define the matrix $X = [x_{ij}]$ $(i,j = 1, 2, \ldots, 6)$ by

$$x_{ij} = \Sigma\, (f_{si} - f_{ti})(f_{sj} - f_{tj}).$$

[all speakers s, t not of the same language]

Define the matrix $T = [t_{ij}]$ $(i,j = 1, 2, \ldots, 6)$ by

$$t_{ij} = \Sigma\, (f_{si} - f_{ti})(f_{sj} - f_{tj})$$

[all speakers s, t]

The columns of the desired transformation V are the eigenvectors of the matrix $T^{-1}X$.

Application of the transformation V to the data space and classification using the nearest-mean decision rule resulted in correct classification of 36 of the 50 training speakers (72%).

Each component of each data vector is a count of the number of occurrences of one sound type for one speaker during a fixed processing time. This count varies directly with the particular rate of speech of the speaker being processed. In an attempt to use features with less dependence on individual speaking rate, ratios of the original components were considered as new features. Of the 15 possible ratios, the six ratios that provided the lowest language uncertainty were used as new features. They are: 2/3, 2/4, 2/5, 2/6, 3/5, and 4/6, where i/j means the ratio of class i to class j. Transformation of the data space by the corresponding transformation V and application of the nearest-mean decision rule to the training speakers yielded 32 correctly classified (64%).

Experiments were done to determine the best subset of the six original sound class features for classifying the training speakers. Consideration of all 63 nonempty subsets showed the use of only the four sound classes v-f, v-n, f-v, and s-v yielded the most training speakers correctly classified. Forty (80%) training speakers were classified correctly in this case.

3. Application to English Data

In order to more closely observe algorithm operation in the use of sound classes, use was made of 14 analog recordings of English data obtained during the course of a password detection study performed at Texas Instruments.[3] Each recording consists of approximately 200 seconds of text read by the speaker. As indicated in Table 5, each of five different speakers read identical text, while Speaker #5 read a total of ten distinct texts.

Sound class recurrences were determined by processing each of the 14 passages in their entirety. The resultant numbers are shown in Table 5. If the sound class detector had been perfect, the first five recurrence values in each of the last six columns of Table 5 would have been identical, since the same text was used for the first five recordings. In view of the rather large variations actually obtained, detailed study was made of algorithmn operation using computer-printed spectra of the English passages and manual transcriptions of the corresponding speech data. Gross differences were observed from speaker to speaker in the nature of the component transitions of the sound classes. These differences caused excessive errors (both false alarms and missed detections of valid occurrences). It was hypothesized that tailoring component reference transitions to individual speakers could significantly reduce the problem.

4. Speaker-Specific Vowel-Nasal Reference Sounds

The experiments described in this subsection involve the vowel-nasal sound class only. Five new reference files of vowel-nasal transitions

TABLE 5. APPLICATION OF SOUND CLASS ANALYSIS TO ENGLISH DATA BASE

Recording #	Speaker #	Text #	v-f	v-n	v-s	f-v	n-v	s-v ← Sound Class	
			Number of Recurrences Detected						
1	1	1	63	248	100	184	94	203	Same Text
2	2	1	108	170	101	191	35	230	Same Text
3	3	1	122	258	155	243	91	278	Same Text
4	4	1	116	208	161	167	99	194	Same Text
5	5	1	104	187	122	182	97	223	Same Text / Same Speaker
6	5	2	94	228	100	190	104	216	Same Speaker
7	5	3	117	235	105	188	106	226	Same Speaker
8	5	4	98	201	120	195	83	243	Same Speaker
9	5	5	93	222	159	188	96	250	Same Speaker
10	5	6	103	226	142	208	97	252	Same Speaker
11	5	7	84	213	175	199	97	259	Same Speaker
12	5	8	115	211	153	208	101	281	Same Speaker
13	5	9	97	240	176	233	95	274	Same Speaker
14	5	10	123	179	132	197	94	241	Same Speaker

were formed. Each new file contains eight transitions of the same types as in the original file of vowel-nasal transitions. However, the reference patterns within the first new file were extracted from data from Speaker #1 in the English data base. Likewise for the other four files, so that speaker-specific reference files were obtained for the vowel-nasal sound class. All the data from each of the five speakers were processed with the corresponding speaker-specific file. The occurrence acceptance threshold, EMAX, was set to 70 for these experiments. Table 6 compares the set of occurrence values obtained using these speaker-specific files with the set of values obtained previously using the speaker-independent file. Also shown are values for the sample mean μ; the standard deviation σ; and σ/μ, a normalized dispersion measure. The variability was decreased significantly by the use of the speaker-specific files as indicated by the 32% decrease in the dispersion measure.

Detailed study of speech spectra and occurrence times of the scanning valleys for the various transitions used in the above experiments led to the conclusion that the 50 ms duration of the reference scanning transitions was too short to allow reliable occurrence detection. Due to the continuous nature of the speech, transitions occur rapidly and often with little stress, resulting in much variability even within tokens of a single speaker. However, it was hypothesized that the 50 ms sound representation would be sufficient to accurately detect single sounds of a more steady-state nature.

C. Detection of Steady-State Sounds

This section describes initial experiments in the detection and use of single, short, steady-state sounds. The data processing algorithm was modified to prevent detection of multiple scanning valleys that might occur during a single target sound. This modification involved monitoring the spectral transitionitivity T.[2] Should more than one scanning valley occur

TABLE 6. VOWEL-NASAL OCCURRENCE VALUES -- SPEAKER-INDEPENDENT VERSUS SPEAKER-SPECIFIC REFERENCES

	Occurrence Values							
Speaker # →	1	2	3	4	5	μ	σ	σ/μ
Speaker-Independent	248	170	258	208	187	214.2	38.1	0.178
Speaker-Specific	261	231	215	208	190	221.0	26.8	0.121

between times of two consecutive peaks of the T function, the modified algorithm selects the one time corresponding to the lowest scanning error as the time of occurrence of the reference sound.

The numbers of peaks in the T function occurring in the first 120 seconds of nonsilence data from the training speakers of L1, L2, L3, L4, and L5 are shown in Table 7. The lack of variation is striking. Approximately 9.2 T peaks occur each second regardless of speaker and language.

1. Detection of Nasals

Consideration was given to identifying the occurrences of the nasal phonemes in Text #1 as read by each of the five speakers in the English data base. There are 180 nasals (m, n, or ŋ) in the idealized text. One pattern was extracted from the data for each of the five speakers which was typical of his nasal sounds. Each of the five complete readings of Text #1 was scanned with each of the five nasal reference sounds as one-element reference files (the occurrence acceptance threshold in these experiments was EMAX = 80). The numbers of recurrences of each reference nasal in each speaker's data are shown in Table 8. Normalized dispersions were computed for each of the five cases where a single reference scanned all five speakers' data (rows of Table 8), and also for the speaker-specific case (diagonal of Table 8). These values are shown in Table 9. It should be noted that the speaker-specific dispersion (0.127) is 40% lower than even the lowest dispersion (0.213) obtained from use of a single reference to scan all five passages.

2. Detection of Sibilants

The experiment in the preceding subsection was repeated to study the detection of the sibilant /s/ (see). Corresponding resultant occurrence values and dispersion values are shown in Tables 10 and 11, respectively. It can be seen that data from Speaker #1 yielded consistently fewer detected sibilants /s/ than data from the other four speakers. It was observed that

TABLE 7. NUMBER OF T PEAKS IN 120 SECONDS OF SPEECH DATA

Language	L1	L2	L3	L4	L5
Speaker					
1	1129	1137	1158	912	1013
2	1085	1195	1167	1125	1148
3	1151	1081	1179	843	1023
4	1161	1167	1147	1065	1180
5	1170	1173	1191	1041	1093
6	1102	1202	1092	1037	1167
7	1060	1117	1156	1117	1085
8	1126	1153	1175	1114	1128
9	1023	1132	1229	1101	1254
10	1072	1149	1055	1042	1275
Mean	1108	1151	1155	1040	1137

TABLE 8. OCCURRENCES OF NASAL SOUNDS

		Source of Data Processed				
	Speaker # →	1	2	3	4	5
Source of Reference Pattern ↓	1	194	5	19	160	135
	2	175	138	193	151	109
	3	307	85	177	235	181
	4	228	9	31	167	142
	5	242	7	49	193	156

TABLE 9. DISPERSION MEASURES FOR NASAL SOUNDS

	Source of Reference Pattern					Speaker-Specific
Speaker # →	1	2	3	4	5	
Std. dev., σ	85.5	32.6	81.8	92.9	98.6	21.1
Mean, μ	102.6	153.3	197.0	115.4	129.4	166.4
σ/μ	0.833	0.213	0.415	0.805	0.762	0.127

TABLE 10. OCCURRENCES OF SIBILANT /s/

Source of Reference Pattern	Speaker #	Source of Data Processed: 1	2	3	4	5
	1	86	102	106	137	118
	2	75	96	103	123	100
	3	82	98	104	134	112
	4	82	96	104	131	106
	5	86	110	121	141	112

TABLE 11. DISPERSION MEASURES FOR SIBILANT /s/

Speaker # →	Source of Reference Pattern 1	2	3	4	5	Speaker-Specific
Std. Dev., σ	19.0	17.2	19.1	17.9	19.9	17.1
Mean, μ	109.8	99.4	106.0	103.8	114.0	105.8
σ/μ	0.173	0.173	0.180	0.172	0.174	0.161

many /s/ sounds from this speaker were accompanied by a "whistle," and most of these whistled sibilants were not detected. Also, the sibilant /ʃ/ (show) in the data for Speaker #4 was similar enough to /s/ so that this phoneme was often detected, accounting for the typically higher values obtained for Speaker #4.

A useful observation was made in the course of determining reference sounds used in the experiments described in the previous two subsections. This observation concerns the choice of extraction time for specific reference patterns. It was seen that the most accurate recurrence detection of steady-state sounds was obtained by using reference sounds extracted at valley points of the T function.

3. Speaker Adaptation

As illustrated in the previous two subsections, accurate recurrence detectability increases when reference sounds are used which reflect the individual speaker's spectral characteristics. An iterative method of speaker adaptation of reference sounds was implemented. This procedure attacks the speaker variability problem by allowing a generalized reference scanning pattern to be modified slightly to accommodate the particular spectral characteristics of the speaker being processed. A small amount of the speaker's data is processed to perform reference pattern modification. We call this "adaptation data." Then the adapted pattern is used for processing the bulk of the data to extract classification parameters. The details of the procedure follow.

Let V_0 denote the given scanning pattern. The first step in the iterative procedure is to scan the adaptation data with V_0, extracting patterns at each significant (low-error) scanning valley. Form V_1, the first modified pattern by letting V_1 be the average of all such extracted patterns. Note $e(V_1,V_0)$, the squared error between V_1 and V_0.

Next re-scan the adaptation data with V_1, again extracting patterns at scanning valleys. Let V_2 denote the average of the pattern extracted using V_1. Note $e(V_2, V_1)$. Repeat this procedure until $e(V_i, V_{i-1})$ is less than or equal to a predetermined threshold. If the threshold is first passed at step N, then V_N is the final reference pattern.

This procedure was applied to Text #1 data as read by the five speakers in the English data base. The beginning scanning pattern in each of the five trials was a nasal /n/ extracted from data from a sixth L6 speaker. The adaptation data comprised the first 75 seconds of speech data from each speaker. The cut-off threshold for each trial was 3.0. That is, the procedure terminated when the squared error between the newest pattern and the previous pattern became 3.0 or less. The resultant five speaker-specific patterns, the beginning pattern V_0, and the five hand-picked patterns described above (Section III.C.1) are all shown in Figure 3.

All Text #1 data from each of the five speakers were then processed to determine nasal occurrences using the corresponding speaker-specific pattern as the reference. The five resultant occurrence values are shown in Table 12 along with the dispersion measure. This dispersion value (0.113) is slightly lower even than that (0.127) obtained using the five hand-picked nasal references described above (Section III.C.1).

4. Language Classification Comparison -- Speaker-Independent Versus Speaker-Specific References

To try to establish whether speaker adaptation of reference sounds would be useful for language classification, the 50 training speakers of L1, L2, L3, L4, and L5 were classified using two sets of measurements. The first set comprised occurrence values for /n/ and /s/ as determined by using a single /n/ reference sound and a single /s/ reference sound. The two references were extracted from data from a speaker of L6.

Beginning Pattern, V_0

```
($    "B=" )($ $)
($    +B=" )($ B)
(====0====)(==+)
```

L6 Speaker	Adapted Pattern	Hand-Selected Pattern

1	`($    +0=, )($ =)`
	`($    +0=, )($ =)`
	`(=========)(===)`
		`($    +BO, )($ $)`
		`($    +00, )($ $)`
		`(=====+===)(===)`

2	`(00, "+=" )(0'''')`
	`(00, "+=" )(0",)`
	`(=========)(===)`
		`(B=  "+== )($0+)`
		`(B=  "+== )($=B)`
		`(=========)(=+$)`

3	`(0=  "+=" )($"+)`
	`(B+  "+=" )($'''')`
	`(=+=======)(==+)`
		`(B"  '''=+ )($=0)`
		`($"  ++=+ )($,0)`
		`(0===00===)(=,=)`

4	`($    +==" )($,+)`
	`($    +==" )($,+)`
	`(=========)(===)`
		`($    +00, )($ $)`
		`($    =0=, )($ B)`
		`(====0=+==)(==+)`

5	`($    +0=, )(B +)`
	`($    +0=, )(B +)`
	`(=========)(===)`
		`($    =00, )($ 0)`
		`($    =00, )($ B)`
		`(=========)(==0)`

Legend

Symbol		,	"	+	=	0	B	$
Level	0	1	2	3	4	5	6	7

Figure 3 Speaker Adaptation Example

TABLE 12. NASAL OCCURRENCES FROM ADAPTED REFERENCES

L6 Speaker	1	2	3	4	5	σ	μ	σ/μ
No. of Occurrences	212	171	184	197	159	20.9	184.6	0.113

The second set of measurements resulted from the use of 50 speaker-specific references for /n/ and /s/ selected from the particular speaker's data which they will subsequently scan to determine occurrence value measurements. These speaker-specific references were determined by first scanning each speaker's data with the L6 reference patterns for /n/ and /s/ and printing a segment of the spectrum surrounding significant (low error) valley point occurrence times. Then visual inspection allowed the choice of a typical speaker-specific reference for /n/ and /s/. Each speaker's data were then processed with their own two references to determine the speaker-specific occurrence values for the two reference sounds.

For these two cases, the nearest-mean decision rule was used to classify the 50 training speakers. Using data generated by the two L6 patterns, 26 of the 50 training speakers were correctly classified. Using speaker-specific references, 28 of 50 were correctly classified.

D. Summary

It was found that detection of steady-state sounds is more appropriate than detection of sound transitions using the 50 ms representation to process continuous speech. While speaker-specific patterns yielded more consistent results in experiments with L6 data than did speaker-independent references, comparison classification experiments using the five-language data base yielded quite similar results.

SECTION IV

LANGUAGE DISCRIMINATION VIA INTERACTIVELY SELECTED REFERENCE SOUNDS

In Section III, experiments were described which provided information useful for providing general guidelines for reference selection. This section describes the selection and use of particular reference sounds for language discrimination. One speaker of each of the languages L1, L2, L3, L4, and L5 was selected at random. The first 90 seconds of speech data from each of these five speakers were used in determining the structure of the reference sounds. Computer-printed digital spectra of these five 90-second passages were generated and then transcribed by hand to associate English phoneme-like labels with appropriate portions of the spectra. The speaker used for each language is shown in Table 13.

A. Postulation of Candidate References

Analog recordings of the five 90-second passages were monitored repeatedly, simultaneously observing the corresponding printed spectra. The ensuing familiarity with the data allowed postulation of several particular steady-state sounds as being language specific. The following procedure was used to obtain reference patterns resulting from such hypotheses.

First, a scanning pattern was extracted at the approximate center of the proposed steady-state sound. Then the same 90-second passage that yielded the proposed sound was scanned with the extracted pattern noting the occurrence times of significant scanning valleys. Then the printed spectra were observed to verify that the sounds indicated by valley points were indeed versions of the target sound. An average pattern was formed using data from all such verified valley points. Finally, the process was repeated using this new average pattern. Iteration continues until little or no change in valley point times results from the current iteration. The resultant average pattern is the proposed candidate reference pattern. For some of the proposed reference sounds, the extraction time of the initial scanning patterns required

TABLE 13. TRANSCRIBED DATA

Language	L1	L2	L3	L4	L5
Speaker	S4	S4	S2	S7	S4

slight modification to provide stable recurrence detection. A few proposed sounds did not yield enough recurrences to be of use and were abandoned immediately. Fifty candidate reference patterns were selected using this interactive procedure. Four additional scanning patterns (one for each of the four phonemes /r/, /ℓ/, /s/, /n/) were selected in a similar fashion using data from Speaker #5 of the English data base.

B. Selection of Language-Specific References

The 54 candidate references were used to process 180 minutes of speech data from each of the 50 training speakers. The numbers of reference occurrences based on occurrence acceptance threshold EMAX = 50 were recorded. Six references yielded no more than ten occurrences per speaker (averaged over each language) from any of the five languages. These references were abandoned because of their rarity in the data. Twelve references produced a fairly uniform distribution of occurrences over all five languages. These references were abandoned because of their obvious lack of language specificity. The 36 retained candidate reference patterns are shown in Appendix A.

Also recorded during processing of the data was the fraction of speech samples that were taken during periods of silence. This fraction (multiplied by 1000) and the occurrence values for the 36 retained candidate references are shown in Table 14. The silence measure is labeled Feature #20.

For later used in classifying the 50 test speakers, the 37 references were used to process 180 seconds of speech from these test speakers using the same occurrence acceptance threshold (EMAX = 50) used with the training data. Occurrence values and values of the silence measure obtained for the 50 test speakers are shown in Table 15.

The nearest-mean decision rule was used to classify the 50 training speakers utilizing various subsets of the collection of 37 candidate references. The

TABLE 14. CANDIDATE REFERENCE OCCURRENCE VALUES -- TRAINING SPEAKERS

		FEATURE #													
LNG	SPKR	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	16	8	15	7	69	136	19	12	39	36	26	199	13	76
1	2	46	44	19	72	86	82	3	22	42	18	34	238	78	84
1	3	24	6	11	18	19	39	4	10	63	6	16	357	25	17
1	4	43	17	24	19	14	37	4	43	68	36	24	217	51	21
1	5	9	23	7	56	79	68	14	2	41	26	12	233	52	80
1	6	2	6	0	39	122	153	3	0	25	23	43	93	65	130
1	7	3	5	1	1	44	87	7	4	27	14	9	182	15	51
1	8	17	14	6	30	153	160	25	0	25	27	33	128	44	156
1	9	27	71	5	94	105	86	24	6	9	9	18	162	112	111
1	10	60	4	4	8	16	36	29	3	1	3	8	286	14	14
2	1	23	3	0	4	83	102	17	0	200	27	9	386	5	88
2	2	61	6	20	14	10	35	24	2	206	25	5	191	20	11
2	3	30	63	32	30	58	71	33	13	147	22	0	238	6	64
2	4	30	4	4	30	49	71	47	1	139	8	24	243	78	54
2	5	1	22	0	19	92	92	11	0	146	31	6	33	17	92
2	6	34	2	1	13	38	69	27	1	257	40	26	134	24	44
2	7	34	6	8	2	15	24	19	4	208	31	6	314	5	20
2	8	16	35	3	89	69	80	28	17	198	48	36	170	80	73
2	9	144	2	28	3	0	19	7	12	179	61	3	211	3	0
2	10	37	5	22	8	28	62	25	22	164	35	8	206	4	30
3	1	10	25	20	14	80	93	1	29	147	114	22	40	42	79
3	2	62	47	33	38	89	82	7	25	118	50	18	191	33	89
3	3	9	19	24	3	39	32	0	74	89	70	3	16	10	41
3	4	11	18	12	41	91	126	4	18	199	77	26	108	73	96
3	5	19	35	15	13	85	100	13	5	116	75	15	80	16	96
3	6	24	19	4	35	16	17	2	10	132	36	10	276	15	16
3	7	3	21	10	9	41	66	3	9	183	73	5	57	15	45
3	8	14	22	11	17	40	76	1	48	208	68	5	89	9	50
3	9	29	76	51	72	75	84	4	114	158	75	16	100	75	85
3	10	11	5	0	0	19	42	1	5	113	45	2	226	0	20
4	1	80	29	24	28	2	0	0	87	24	2	0	417	0	2
4	2	14	7	10	22	11	11	2	39	92	9	2	73	19	11
4	3	61	3	12	25	3	3	1	52	112	28	1	341	14	4
4	4	118	33	22	44	11	6	1	38	74	3	1	300	17	11
4	5	22	28	41	42	44	46	9	24	76	4	10	361	26	45
4	6	84	40	40	17	6	5	2	57	81	9	1	364	2	7
4	7	51	12	49	80	8	14	1	89	27	2	2	374	51	9
4	8	42	32	32	65	8	5	4	17	175	40	1	215	43	9
4	9	53	60	43	66	20	8	8	42	93	4	2	292	10	18
4	10	77	30	53	72	17	11	10	33	90	11	3	226	54	17
5	1	20	37	12	34	56	56	2	25	147	2	11	366	9	56
5	2	47	37	40	55	34	35	7	20	132	7	13	283	32	32
5	3	78	53	28	76	111	83	0	65	59	3	5	359	22	114
5	4	79	159	60	88	51	20	10	50	98	4	2	291	21	54
5	5	44	57	21	71	22	13	5	34	148	9	5	404	13	18
5	6	88	59	16	76	26	37	5	95	122	3	8	238	32	27
5	7	58	19	6	136	67	60	3	35	110	1	7	335	56	70
5	8	38	62	29	80	85	86	5	45	64	8	11	346	24	90
5	9	13	79	13	97	54	42	12	5	198	52	3	30	73	64
5	10	0	9	3	5	27	20	2	9	281	151	0	48	1	29

TABLE 14
(Continued)

FEATURE #

LNG	SPKR	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	1	72	156	6	2	0	240	0	205	5	3	112	10	4	1
1	2	212	250	10	10	38	296	6	276	29	17	259	42	40	24
1	3	154	340	7	2	4	278	3	372	13	4	237	27	3	3
1	4	120	204	15	9	8	221	19	222	57	11	170	43	10	8
1	5	50	198	2	1	23	363	0	236	2	6	88	1	29	15
1	6	22	60	0	0	17	435	0	76	0	0	22	0	22	2
1	7	35	102	0	0	0	256	0	131	2	0	49	5	7	1
1	8	52	77	5	0	11	282	2	119	10	2	51	1	14	5
1	9	104	181	13	5	54	389	0	165	1	14	119	7	59	35
1	10	148	280	1	0	5	401	4	289	4	0	173	14	1	1
2	1	212	375	0	0	2	158	1	378	1	1	217	5	9	1
2	2	115	131	9	2	10	86	9	201	31	5	105	13	1	3
2	3	156	245	22	10	15	172	7	264	21	27	196	15	40	42
2	4	103	176	1	0	15	218	1	234	2	1	104	2	10	2
2	5	4	24	0	0	7	263	0	26	1	0	4	0	25	5
2	6	64	77	0	0	7	105	5	127	8	1	71	2	1	1
2	7	207	285	1	0	1	179	8	319	15	0	229	10	3	1
2	8	43	135	6	2	64	268	4	164	8	4	65	11	31	23
2	9	165	224	7	6	3	229	36	240	78	7	189	26	0	1
2	10	67	141	5	2	4	205	1	188	32	3	94	12	2	3
3	1	14	44	6	8	7	347	0	38	10	5	41	18	30	12
3	2	119	146	16	10	26	383	10	215	46	19	130	13	54	34
3	3	32	32	2	5	0	348	1	24	17	2	58	36	19	3
3	4	50	118	8	4	22	277	0	132	23	3	72	9	25	7
3	5	32	58	14	2	8	500	1	76	5	20	44	7	33	11
3	6	93	187	4	8	16	269	7	250	18	8	108	17	9	11
3	7	9	35	8	4	6	300	0	28	2	2	14	1	14	4
3	8	29	73	18	32	3	355	0	88	17	22	56	25	11	19
3	9	64	114	35	28	37	319	7	101	86	36	112	51	58	50
3	10	43	206	1	1	0	173	1	219	4	1	55	3	4	3
4	1	344	412	25	46	16	281	24	464	57	35	360	72	7	26
4	2	67	77	0	12	2	251	0	93	28	4	86	11	2	3
4	3	258	333	1	7	8	323	4	342	65	7	283	59	1	4
4	4	349	332	23	19	34	226	28	364	32	39	386	83	19	33
4	5	183	286	26	14	21	185	4	354	15	11	180	13	25	16
4	6	266	352	44	44	11	229	13	376	56	45	317	59	18	33
4	7	265	320	25	32	54	142	22	408	93	20	303	21	1	17
4	8	112	168	32	25	29	141	14	195	24	25	116	20	12	24
4	9	123	178	42	44	52	248	18	266	39	43	116	18	27	50
4	10	249	212	32	17	50	241	22	254	54	23	272	38	15	29
5	1	106	315	13	15	36	378	10	345	37	19	141	21	33	33
5	2	98	226	28	19	33	237	14	297	57	24	153	16	25	18
5	3	204	342	15	21	36	226	9	397	51	13	251	42	52	36
5	4	258	363	108	72	69	261	24	386	64	141	382	52	88	131
5	5	363	396	39	43	57	40	22	439	32	60	400	38	31	60
5	6	167	250	33	64	45	248	35	283	60	49	225	80	28	50
5	7	216	291	1	15	78	101	13	368	32	12	258	32	24	20
5	8	172	304	41	34	58	239	13	374	69	55	218	31	57	52
5	9	26	18	21	8	45	205	0	29	3	38	32	3	68	66
5	10	10	51	1	0	1	182	1	37	1	0	15	5	9	0

TABLE 14
(Continued)

LNG	SPKR	29	30	31	32	33	34	35	36	37
		FEATURE #								
1	1	150	2	115	73	0	6	144	228	14
1	2	290	14	260	230	0	9	284	322	18
1	3	274	8	197	157	1	10	271	366	10
1	4	181	34	149	158	1	16	161	213	49
1	5	141	2	100	29	6	2	171	240	3
1	6	38	0	34	14	3	0	41	54	0
1	7	69	1	56	24	0	1	78	118	2
1	8	85	13	66	72	13	5	81	113	3
1	9	145	3	119	100	1	3	152	176	4
1	10	221	7	166	152	0	4	239	270	4
2	1	288	2	259	167	13	0	308	388	0
2	2	153	40	154	146	5	18	122	212	15
2	3	217	5	174	181	3	12	200	266	15
2	4	149	5	136	120	9	1	134	225	1
2	5	11	0	5	2	2	0	11	17	0
2	6	91	12	84	90	1	2	71	124	3
2	7	285	16	252	196	0	4	273	336	2
2	8	79	3	54	61	15	2	85	129	4
2	9	220	105	200	189	9	19	212	241	30
2	10	109	21	93	91	12	9	94	153	27
3	1	31	11	23	20	33	12	38	45	20
3	2	190	34	164	117	7	10	153	187	27
3	3	35	12	44	49	19	15	38	53	59
3	4	92	13	73	50	18	7	97	140	8
3	5	54	9	44	38	5	4	50	64	5
3	6	115	16	121	89	8	6	103	225	11
3	7	15	0	14	9	8	3	23	32	8
3	8	63	5	43	54	3	1	52	75	22
3	9	114	48	95	90	4	40	119	138	93
3	10	98	9	72	14	6	0	122	186	5
4	1	436	59	412	354	8	13	375	485	55
4	2	109	12	77	88	29	7	90	105	34
4	3	325	38	298	238	17	18	290	368	57
4	4	417	34	399	390	28	13	367	417	44
4	5	269	19	235	168	15	24	220	350	38
4	6	354	55	329	269	14	26	328	427	75
4	7	365	89	348	259	21	45	306	419	99
4	8	161	38	134	113	1	27	143	182	34
4	9	140	45	145	125	16	19	93	243	57
4	10	286	56	275	268	1	25	244	290	46
5	1	177	21	143	105	4	9	142	254	20
5	2	193	36	147	154	12	11	143	228	19
5	3	297	18	256	232	7	17	262	370	48
5	4	442	80	302	318	11	20	383	372	44
5	5	439	46	407	360	1	9	399	439	36
5	6	266	49	210	248	13	4	221	284	46
5	7	319	31	270	262	2	4	275	364	20
5	8	246	57	222	177	13	13	220	324	31
5	9	35	10	29	31	2	4	23	27	5
5	10	18	1	14	12	2	1	36	37	6

TABLE 15. CANDIDATE REFERENCE OCCURRENCE VALUES -- TEST SPEAKERS

		FEATURE #													
LNG	SPKR	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	5	10	15	42	125	123	13	6	20	6	54	185	83	125
1	2	4	1	11	2	17	30	0	12	40	53	0	212	6	19
1	3	11	19	14	27	25	52	1	11	33	16	7	128	19	30
1	4	5	10	29	20	22	52	2	20	40	32	1	104	10	24
1	5	9	17	20	44	163	168	22	3	5	18	26	139	49	164
1	6	18	12	12	104	27	25	1	7	23	16	9	299	105	26
1	7	12	18	48	28	11	15	1	21	12	14	1	160	11	11
1	8	16	14	14	39	11	13	1	8	54	34	1	237	30	11
1	9	21	14	18	20	29	29	4	22	32	50	2	159	5	31
1	10	15	5	18	2	44	68	8	9	24	14	2	196	0	46
2	1	7	36	25	32	82	113	5	4	122	35	17	132	23	88
2	2	9	8	50	74	12	16	2	14	158	31	12	236	37	14
2	3	13	4	52	96	48	60	15	13	71	17	20	193	97	51
2	4	23	89	75	88	22	3	0	40	123	18	1	247	2	15
2	5	17	25	14	56	21	19	7	0	130	33	4	263	26	23
2	6	21	9	17	15	27	22	22	5	132	34	1	250	7	26
3	1	7	18	18	1	39	62	1	14	205	77	10	138	2	40
3	2	8	9	7	3	40	59	0	13	210	64	6	87	10	43
3	3	0	9	20	6	78	83	0	15	211	70	8	14	9	78
3	4	5	13	26	6	60	84	12	15	139	57	19	36	17	65
3	5	0	24	1	60	122	104	0	3	175	69	7	0	114	140
3	6	2	34	6	15	128	110	0	25	241	87	5	35	20	132
3	7	9	32	29	39	155	128	6	11	175	51	15	91	34	156
3	8	5	28	41	17	62	49	0	44	236	40	2	75	17	63
3	9	11	17	16	23	10	7	0	25	105	20	0	233	1	9
3	10	2	34	11	59	41	28	0	70	124	50	3	43	11	40
4	1	6	1	1	15	44	68	27	0	95	46	21	266	18	48
4	2	4	32	49	62	12	9	1	43	120	7	0	278	9	12
4	3	35	66	70	90	68	26	39	10	146	2	1	206	37	61
4	4	14	2	9	28	1	0	1	0	134	13	0	320	11	1
4	5	9	1	5	3	6	24	8	3	74	8	4	409	2	4
4	6	17	6	21	43	31	30	9	19	61	21	11	344	52	35
4	7	47	3	20	9	5	6	12	7	105	7	0	304	2	6
4	8	10	43	83	33	44	22	28	12	92	45	2	268	8	37
4	9	16	9	56	65	2	0	0	35	130	20	0	354	27	2
4	10	27	31	60	22	7	11	3	17	83	13	0	145	13	8
4	11	21	5	12	51	0	0	0	10	91	6	0	297	4	0
4	12	15	26	29	84	73	64	17	5	22	18	14	287	68	76
4	13	13	8	57	14	8	20	12	8	105	18	2	98	14	12
4	14	6	48	7	161	33	0	0	160	3	0	0	294	10	23
5	1	5	17	1	30	61	76	2	2	222	66	1	69	68	68
5	2	18	58	8	200	27	3	0	27	141	27	0	181	108	27
5	3	9	34	27	41	31	37	2	6	152	32	4	100	16	33
5	4	19	50	10	73	36	10	0	18	191	110	0	117	10	34
5	5	10	34	25	39	36	34	3	8	174	59	4	73	17	34
5	6	22	36	44	20	24	15	4	15	167	70	1	72	4	24
5	7	3	38	13	34	109	87	0	34	148	39	3	129	14	111
5	8	15	6	6	12	48	38	0	9	227	63	2	126	5	49
5	9	30	25	14	64	47	26	0	7	119	104	5	179	37	48
5	10	15	13	21	8	11	40	0	10	246	58	2	99	12	14

TABLE 15
(Continued)

LNG	SPKR	15	16	17	18	19	20	21	22	23	24	25	26	27	28
		FEATURE #													
1	1	36	157	6	5	25	372	2	177	17	3	63	9	24	12
1	2	93	196	2	6	0	233	3	213	31	4	120	17	3	0
1	3	84	149	13	16	11	211	3	138	11	20	125	15	16	19
1	4	134	179	22	23	6	184	1	161	40	19	191	39	8	11
1	5	73	98	10	3	20	223	1	145	14	6	82	8	42	18
1	6	207	267	10	10	48	297	7	343	20	15	248	22	14	16
1	7	166	199	27	53	16	243	33	205	90	46	201	49	11	19
1	8	122	284	14	23	15	348	20	269	62	29	182	21	10	15
1	9	116	158	17	34	12	314	12	170	32	25	178	61	12	20
1	10	111	250	9	7	2	275	13	228	31	4	205	16	8	2
2	1	44	75	33	17	16	306	2	98	12	17	45	8	27	30
2	2	101	227	12	28	52	127	34	240	99	31	154	12	7	25
2	3	93	160	16	6	65	266	23	197	82	7	143	34	10	13
2	4	127	195	81	108	36	214	16	248	48	99	168	25	54	79
2	5	90	186	22	25	23	314	3	217	10	56	118	10	16	45
2	6	193	270	17	30	8	322	9	278	69	26	265	51	14	11
3	1	52	98	12	8	1	258	0	135	14	4	63	20	22	12
3	2	14	58	3	2	2	232	1	73	5	5	25	18	11	9
3	3	6	32	6	6	3	229	0	20	10	2	29	14	28	7
3	4	20	23	19	4	1	309	0	25	6	5	39	17	17	11
3	5	0	0	0	1	14	343	0	0	2	2	0	3	62	21
3	6	14	33	2	3	6	254	0	29	11	1	48	37	65	12
3	7	15	28	16	2	14	384	1	52	11	6	13	2	77	23
3	8	29	79	27	27	8	260	4	86	79	14	55	23	33	18
3	9	103	108	9	21	8	258	7	190	45	6	97	35	10	12
3	10	17	24	18	35	19	352	2	35	26	22	27	23	24	31
4	1	97	209	0	1	7	188	2	251	1	2	98	0	3	2
4	2	136	269	40	89	27	277	14	283	57	66	176	42	23	37
4	3	109	177	42	17	50	249	37	212	86	25	104	22	57	64
4	4	47	182	3	13	18	272	17	243	34	15	60	3	2	11
4	5	203	400	3	1	1	374	2	399	6	0	218	13	2	1
4	6	109	241	3	3	26	203	6	323	42	2	108	10	12	8
4	7	129	230	5	6	3	226	16	302	57	9	130	13	2	8
4	8	132	214	50	29	21	114	16	267	55	39	141	14	42	36
4	9	238	327	31	79	43	260	49	371	120	53	258	29	2	27
4	10	176	147	41	41	10	307	9	196	60	48	179	13	16	39
4	11	193	199	14	65	26	201	15	290	38	22	205	17	0	25
4	12	121	226	21	10	49	221	7	272	13	16	150	7	31	26
4	13	108	64	22	11	3	179	8	100	35	12	102	8	3	13
4	14	336	320	35	135	96	53	28	350	38	41	372	129	34	67
5	1	29	33	3	1	8	232	1	63	10	2	27	2	28	9
5	2	71	137	11	35	129	322	14	172	37	39	80	19	37	107
5	3	30	39	25	11	25	264	7	66	35	36	39	11	39	40
5	4	60	41	8	13	27	321	0	95	28	26	66	21	43	60
5	5	20	26	24	11	20	256	6	41	21	30	30	12	38	41
5	6	34	40	31	17	7	196	2	54	29	20	29	6	33	30
5	7	46	93	16	15	15	185	3	133	17	13	72	25	73	35
5	8	60	50	3	0	4	265	2	103	8	1	54	12	16	4
5	9	131	110	11	4	20	213	7	196	31	8	134	13	27	25
5	10	31	20	18	6	0	247	7	67	17	10	36	14	6	8

TABLE 15
(Continued)

		FEATURE #								
LNG	SPKR	29	30	31	32	33	34	35	36	37
1	1	102	18	49	42	0	8	119	156	17
1	2	158	21	121	116	0	10	162	208	13
1	3	135	9	102	111	0	4	146	156	8
1	4	185	17	154	168	0	18	178	195	37
1	5	108	9	96	108	0	7	98	131	10
1	6	318	13	246	240	1	5	264	324	9
1	7	234	76	209	205	0	22	213	257	75
1	8	226	59	160	143	1	15	266	298	31
1	9	177	34	134	172	1	8	144	191	39
1	10	238	27	157	141	1	11	253	242	17
2	1	58	18	59	47	9	5	56	102	17
2	2	193	81	136	125	45	24	189	227	62
2	3	156	43	116	125	0	20	139	195	40
2	4	209	65	161	140	9	35	186	235	81
2	5	125	21	122	94	10	10	130	197	16
2	6	287	41	222	245	13	16	268	333	43
3	1	91	4	76	63	12	8	70	114	27
3	2	33	10	25	21	7	14	34	50	16
3	3	16	2	11	10	23	12	24	30	24
3	4	33	7	27	36	19	14	26	29	18
3	5	0	1	0	0	23	0	0	0	0
3	6	45	6	31	38	19	7	46	50	34
3	7	24	10	25	23	5	15	11	43	15
3	8	66	32	40	56	1	27	68	71	74
3	9	114	16	126	102	6	16	89	139	46
3	10	27	18	23	35	21	15	26	28	67
4	1	135	2	114	79	9	0	136	221	0
4	2	218	39	165	112	21	25	227	293	85
4	3	135	80	135	133	31	45	117	190	66
4	4	101	28	71	55	10	3	83	184	11
4	5	289	8	251	130	16	7	313	405	8
4	6	196	31	153	87	24	14	172	307	41
4	7	217	49	167	142	1	19	175	266	27
4	8	172	60	155	122	10	48	144	259	63
4	9	300	132	278	187	21	31	288	386	117
4	10	197	56	210	197	42	35	164	214	66
4	11	240	35	220	196	6	6	198	277	19
4	12	201	26	159	105	27	17	186	260	22
4	13	109	46	135	124	0	35	80	133	48
4	14	379	27	369	362	0	2	341	385	79
5	1	33	10	39	33	3	2	24	51	3
5	2	99	33	90	90	10	4	76	153	22
5	3	48	28	37	39	20	9	25	43	22
5	4	83	11	72	82	17	7	48	75	24
5	5	33	22	24	28	18	11	22	33	18
5	6	46	25	38	44	5	20	27	41	33
5	7	83	15	53	63	5	5	68	87	26
5	8	74	2	81	90	34	3	42	72	5
5	9	158	27	162	167	35	13	113	168	12
5	10	36	25	35	45	12	14	23	39	26

following semi-automatic procedure was used to establish which collection of reference subsets to use for classification. A subset of N, $N \leq 12$, of the 37 references was selected. Classification was performed for each of the $2^N - 1$ non-empty subsets of the collection of N initial references, examining the results by eye to find those subsets that allow the maximum number of correct decisions. Those references that seem not to contribute to good classification were discarded and replaced by others in the set of 37-N references not used initially. Another exhaustive search was made using the new collection of references, and the procedure was then iterated until all 37 references had been considered. A variation of this procedure allowed consideration of larger subsets, where the exhaustive search was not feasible. For such subsets, M component references (which have previously been deemed good for language discrimination) were included in every subset evaluated. Thus, only $2^{N-M} - 1$ evaluations were required.

The maximum number of correctly classified training speakers was 47 (94%). The five subsets of the 37 original references which provided this classification accuracy are shown in Table 16.

The 50 test speakers were classified using each of the five subsets with the nearest-mean decision rule, based on mean values obtained from the training data results for the corresponding subset. For each of the five subsets, the percent correct classification of the test speakers is also shown in Table 16. The maximum test data accuracy was 66%, attained using the subset of 14 references. Appendix B contains listings of the decision function values and resultant classifications for training speakers and test speakers, for all five subsets of references.

TABLE 16. SUBSETS OF CANDIDATE REFERENCES YIELDING 94% *CORRECT CLASSIFICATION -- TRAINING SPEAKERS*

1 (6 Refs.)	2 (7 Refs.)	3 (12 Refs.)	4 (14 Refs.)	5 (16 Refs.)
6	2	6	6	6
9	7	7	7	7
10	9	9	8	8
11	10	10	9	9
19	20	11	10	10
24	27	20	13	11
	34	21	19	13
		24	20	17
		27	22	20
		28	24	21
		33	27	22
		34	28	24
			33	27
			34	28
				33
				34
42%	60%	62%	66%	60%

Percent Correct Classification -- Test Speakers

C. Long-Term Spectral Normalization

The test speaker classification accuracy (66%) attained as described in the previous section was disappointingly low in view of the rather high language specificity exhibited by the references as determined from the training data. This poor performance prompted close scrutiny of the test speaker occurrence results (Table 15) as compared with corresponding values for the training speakers (Table 14). As a result, it was noticed that, as a group, data from training Speakers #1 through #8, language L5, were strikingly different from the group of data comprising training Speakers #9 and #10 and test Speakers #1 through #10, all from language L5. For example, the occurrence values for Reference #10 are all less than 10 for the first group and larger than 26 for the second group. This observation prompted listening to a speaker from both groups for comparison. It was apparent from this aural comparison that the recording conditions were quite different for the two sessions. Data collection documentation indicated that five different sources of data and three different sets of recording equipment were used to obtain the 100 sessions of data used in the experiments. The sources were: (A) recordings from Defense Language Institute, East Coast; (B) recordings from Defense Language Institute, West Coast; (C) recordings made at the University of Pennsylvania; (D) recordings from Radio Free Europe, New York; and (E) recordings made in two rooms of a church. The three recorder/microphone equipment sets were: (1) NAGRA/D24E, (2) AMPEX/ALTEC, and (3) AMPEX/SHURE 55S. For each speaker, Table 17 shows the source of the data and the recording equipment used, if known. It was noted that the two groups of data described above were in fact recorded under different conditions and, further, that all the speakers within a group were recorded under identical conditions. The first group of speakers was recorded in a church building where the reverberation level was high and the high frequencies somewhat attenuated. The second group of speakers was recorded in an acoustically treated room at the University of Pennsylvania, and the frequency range was much larger than for the first group.

TABLE 17. SOURCES OF ANALOG RECORDINGS

Speaker	L1	L2	L3	L4	L5
	Language				
Training Speakers					
1	A1	D	C3	E2	E2
2	A2	D	C3	E1	E2
3	A2	D	C3	E1	E2
4	A2	D	C3	E2	E2
5	A1	D	C3	E2	E2
6	B	D	C3	E2	E2
7	B	D	C3	E2	E2
8	B	B	C3	A2	E2
9	B	B	C3	A2	C3
10	B	B	C3	A2	C3
Test Speakers					
1	B	B	C3	A1	C3
2	B	B	C3	A2	C3
3	B	B	C3	B	C3
4	B	B	C3	B	C3
5	B	B	C3	B	C3
6	B	B	C3	B	C3
7	B		C3	B	C3
8	B		C3	B	C3
9	B		C3	B	C3
10	B		C3	B	C3
11				B	
12				B	
13				E1	
14				E2	

Sources:

A DLIEC
B DLIWC
C Un. of Penn.
D RFE
E Church

Equipment:

1 NAGRA/D24E
2 AMPEX/ALTEC
3 AMPEX/Shure 55S

It was hypothesized that standardization of each speaker's long-term average spectrum, subsequent normalization of filter outputs to this standard, and subsequent reprocessing of the data would remove much of the variation in reference sound detection resulting from the variation in recording conditions. Spectral standardization was accomplished as follows: Let D_j denote the vector of outputs from the 16-filter analog filter bank at time j for some speaker, let A denote the vector of averaged filter outputs where the average is taken over the M samples of actual speech occurring in 60 seconds in input data. That is,

$$A = \frac{1}{M} \sum_{i=1}^{M} D_{j_i}$$

The vector S of normed spectral averages for the given speaker is defined to be

$$S = \frac{A}{A \cdot \underline{1}}$$

where $A \cdot \underline{1}$ is the dot product of vector A and the unit vector $\underline{1}$.

An idealized long-term spectral average vector, denoted Q, was obtained by averaging the normed spectral average vectors from one speaker of each of the languages L1, L2, L3, L4, and L5. These speakers are the five training speakers used to define the candidate references and are listed in Table 13. For a given speaker, let q_k denote the k-th component of Q and let d_{k_j} denote the k-th component of D_j (the output of the k-th filter at time j). Let s_k denote the k-th component of this speaker's long-term spectral average vector, S. Then $d^*_{k_j}$, the k-th component of the spectrally normalized data vector D^*_j (at time j) is defined to be

$$d^*_{kj} = \frac{q_k}{s_k} \cdot d_{kj}$$

Subsequent to this normalization of the data the new long-term spectral average vectors of all speakers (averaged over the same 60-second period) are identical and equal Q. This standard long-term average spectrum Q and the five spectra averaged to obtain Q are shown in Figure 4.

D. Post-Normalization Processing and Classification Results

For each of the 100 speakers in the data base the spectrally normalized filter output data were preprocessed in exactly the same manner as was the original data. This normalized data was scanned with the same set of 36 references described in Section IV.B. These 36 scanning patterns are exactly those patterns used to process the original data. That is, the reference extraction process was not repeated using normalized data. Had more time been available the patterns most certainly would have been reconstructed using the normalized data. Occurrence values and the silence measure determined from the normalized data are shown in Table 18 (training speakers) and Table 19 (test speakers). It can be seen that the anomaly concerning Reference #10, language L5, described in the previous section has disappeared.

Evaluation of various subsets of the collection of 37 features followed the procedure described in Section IV.B. The set union of the subsets found which yielded 94% correct classification of the 50 training speakers is the collection of 15 features $M = \{1,2,5,7,9,10,11,13,14,20,26,27,30,33,34\}$. A forward search with substitutions (FSS) was utilized to evaluate important subsets of M in a systematic and automatic manner. This procedure determines a k-element subset M_k of M for k = 1, 2, ..., 15 which yields classification accuracy of the 50 training speakers that is higher than the accuracy using any other subset of k elements considered.

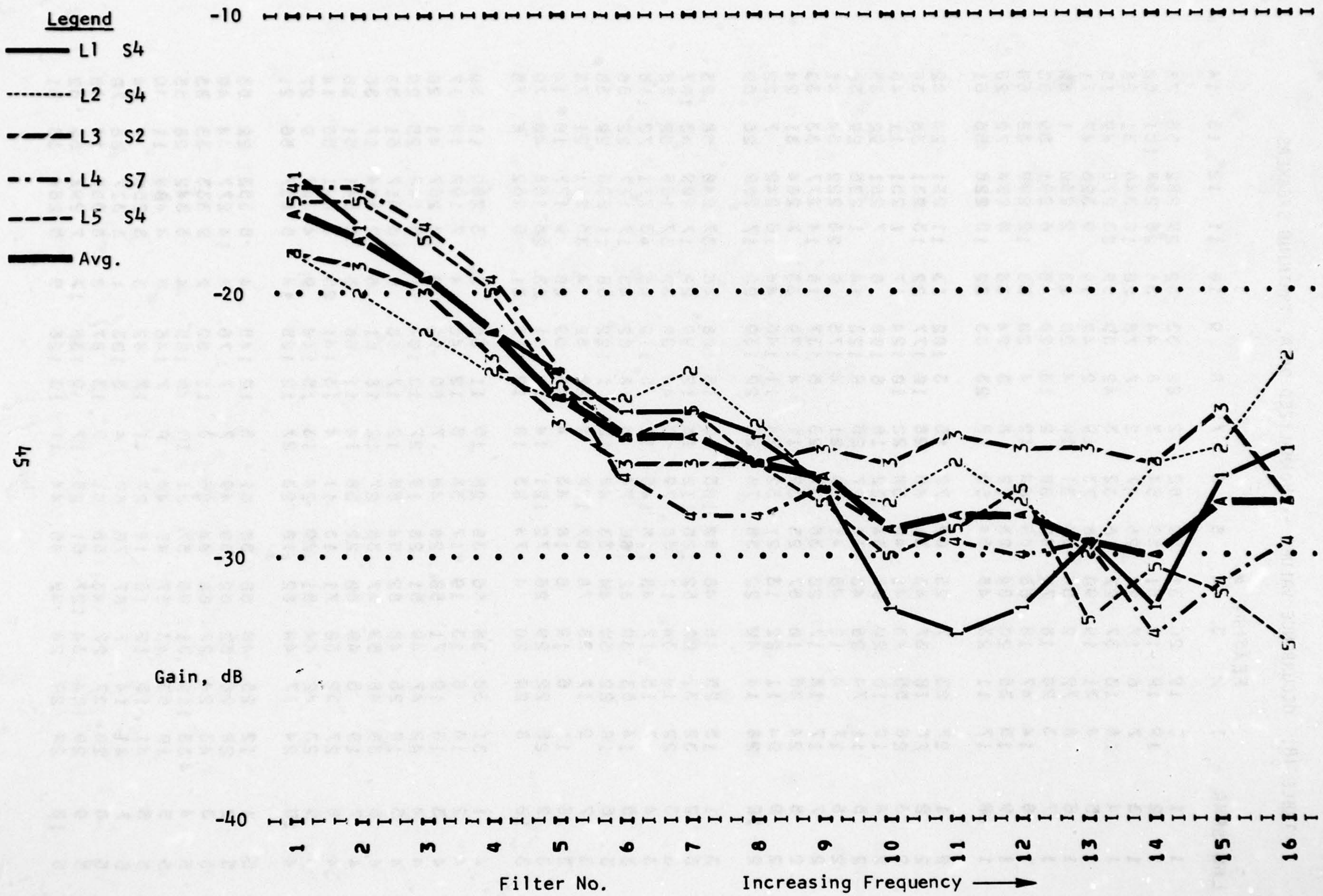


Figure 4 Long-Term Average Spectra

TABLE 18. OCCURRENCE VALUES -- NORMALIZED DATA, TRAINING SPEAKERS

LNG	SPKR	FEATURE # 1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	7	19	26	31	59	82	2	22	33	32	29	282	56	70
1	2	19	19	12	81	56	54	3	8	44	36	26	258	101	60
1	3	7	6	17	27	25	47	2	7	76	28	10	348	31	26
1	4	16	13	37	27	16	32	3	42	59	52	23	275	49	15
1	5	4	31	19	90	68	73	19	2	42	19	9	326	47	71
1	6	6	72	2	56	59	31	10	4	20	23	2	250	1	60
1	7	3	28	18	72	80	80	2	18	29	5	6	291	69	83
1	8	14	47	18	68	66	64	12	4	20	53	12	290	38	65
1	9	15	38	23	84	23	12	8	3	24	36	0	204	72	26
1	10	17	11	23	46	54	53	17	23	33	32	15	226	55	61
2	1	27	23	7	25	82	79	15	3	102	13	11	281	24	82
2	2	26	10	37	43	31	45	28	10	177	22	13	231	38	36
2	3	26	55	43	41	42	50	22	10	124	17	1	251	11	45
2	4	19	15	20	97	34	32	16	6	108	6	7	281	92	35
2	5	11	74	26	43	36	27	25	6	123	14	1	225	29	38
2	6	17	4	12	46	41	59	21	4	175	9	23	222	54	41
2	7	17	14	17	23	36	32	23	5	137	16	14	277	33	33
2	8	24	36	10	67	25	32	11	4	179	33	7	284	31	24
2	9	94	11	22	13	21	52	24	11	146	44	10	242	7	22
2	10	28	14	40	22	58	70	23	20	139	22	17	269	26	59
3	1	15	25	16	48	80	105	0	16	108	46	37	240	58	83
3	2	32	31	22	52	106	112	15	12	99	20	17	209	43	107
3	3	22	19	34	17	68	79	5	44	38	30	37	166	50	70
3	4	7	15	17	48	110	145	12	5	119	46	43	271	72	116
3	5	14	53	39	32	88	74	6	10	82	43	17	177	22	86
3	6	18	22	30	44	33	44	13	1	122	38	11	293	29	35
3	7	9	17	33	78	67	120	2	22	62	4	35	241	91	73
3	8	11	6	19	6	18	43	4	40	69	20	9	177	10	14
3	9	20	22	29	26	72	121	14	14	81	23	26	158	40	79
3	10	8	29	20	4	74	83	10	12	116	31	9	262	6	75
4	1	31	38	36	56	38	28	19	11	56	3	3	289	18	39
4	2	10	6	13	19	17	38	8	12	43	4	7	190	16	17
4	3	13	19	71	58	26	40	7	69	92	9	10	267	41	28
4	4	42	47	49	51	28	19	27	12	104	11	5	146	25	29
4	5	18	28	48	62	54	56	12	17	68	5	16	317	61	53
4	6	38	48	53	47	38	27	12	18	61	6	2	304	17	36
4	7	19	5	49	60	22	30	14	11	66	3	3	341	51	26
4	8	27	32	50	71	13	11	4	13	141	25	3	241	56	14
4	9	23	42	44	51	29	24	13	15	114	9	4	307	9	27
4	10	24	17	40	62	18	23	27	12	105	14	5	257	56	21
5	1	12	25	40	58	58	61	5	12	145	4	6	353	29	63
5	2	28	26	52	62	39	49	7	7	76	4	14	277	44	40
5	3	43	24	27	60	84	86	3	11	69	2	9	333	35	83
5	4	33	119	31	96	57	31	13	5	105	4	5	342	28	55
5	5	10	63	43	57	45	44	9	7	146	5	4	409	11	49
5	6	41	13	12	16	14	27	11	17	89	3	3	256	9	14
5	7	41	14	7	87	70	80	4	3	103	1	3	327	38	78
5	8	20	27	27	43	86	91	9	15	87	8	8	357	16	88
5	9	29	104	34	127	61	28	17	9	136	17	7	297	52	62
5	10	20	29	24	42	45	44	11	13	126	5	5	264	38	51

TABLE 18
(Continued)

LNG	SPKR	FEATURE # 15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	1	165	230	10	4	11	244	4	291	39	4	194	30	29	13
1	2	224	275	13	6	40	288	7	289	22	16	268	43	26	22
1	3	139	320	8	6	9	271	7	342	36	3	193	40	7	7
1	4	124	238	19	16	11	219	21	291	81	14	169	40	13	9
1	5	225	331	8	10	30	370	0	326	14	33	251	1	39	41
1	6	169	268	17	12	20	435	0	288	0	53	186	7	66	78
1	7	219	237	7	3	32	260	0	304	21	12	230	25	37	26
1	8	185	305	13	7	47	292	3	318	29	30	222	7	46	49
1	9	167	245	31	15	51	384	6	220	9	34	184	19	28	42
1	10	120	252	11	6	16	384	5	247	31	6	206	53	14	17
2	1	284	289	7	1	12	179	4	340	26	7	276	19	50	29
2	2	148	179	20	15	19	123	16	250	47	18	162	26	8	13
2	3	181	260	38	17	26	196	11	279	28	46	231	26	48	56
2	4	207	262	8	4	71	207	7	301	26	6	220	5	20	33
2	5	142	202	50	24	23	292	1	230	8	67	152	12	50	63
2	6	193	213	2	3	21	173	8	244	51	7	194	8	10	12
2	7	207	228	9	9	10	176	8	290	16	5	236	20	19	10
2	8	154	281	12	24	65	283	8	299	23	51	203	25	33	42
2	9	164	228	14	15	5	257	56	256	53	12	200	56	10	12
2	10	119	198	14	7	12	224	18	265	52	9	138	24	17	10
3	1	138	206	9	6	24	367	7	229	23	9	145	22	36	20
3	2	152	208	19	7	30	416	8	259	40	14	180	24	47	28
3	3	124	154	10	8	4	382	7	168	19	8	173	70	28	7
3	4	168	271	9	9	24	307	2	293	32	7	189	18	32	8
3	5	118	165	38	19	25	510	6	195	30	35	136	18	65	32
3	6	155	298	17	6	15	292	6	303	35	19	196	3	20	25
3	7	130	244	22	14	56	410	2	237	57	15	150	8	16	21
3	8	110	168	5	21	2	395	17	186	67	9	159	40	4	5
3	9	101	178	17	4	18	369	5	142	35	5	163	51	29	14
3	10	187	332	12	7	1	308	3	319	15	10	252	18	28	22
4	1	143	251	35	19	22	275	10	285	13	38	179	18	31	40
4	2	152	214	8	11	4	252	1	198	17	9	176	10	3	4
4	3	192	281	51	58	40	301	14	283	45	49	254	73	14	36
4	4	148	131	37	24	35	212	9	183	21	42	197	38	40	55
4	5	227	300	41	16	32	202	6	315	32	23	234	10	31	24
4	6	187	283	38	28	33	292	20	326	30	49	227	33	44	60
4	7	188	321	16	15	39	146	14	344	53	10	239	14	6	12
4	8	133	221	40	35	45	181	14	246	34	40	151	19	16	33
4	9	132	256	38	47	40	244	15	284	32	56	136	15	35	55
4	10	250	292	23	9	47	249	13	280	34	14	281	43	15	26
5	1	135	344	24	18	41	330	7	375	53	33	188	8	32	35
5	2	114	290	33	18	39	227	20	293	66	27	161	9	26	28
5	3	191	310	10	11	32	224	15	374	51	7	228	17	24	24
5	4	278	424	84	62	80	239	23	424	45	123	370	40	89	116
5	5	325	438	38	24	46	96	8	449	24	64	401	39	57	71
5	6	216	287	20	42	13	239	21	302	33	21	248	48	11	17
5	7	231	337	1	8	51	126	12	365	50	19	266	15	32	33
5	8	178	354	25	22	33	258	12	387	52	43	230	33	43	38
5	9	210	227	69	58	85	250	2	303	13	97	197	11	78	122
5	10	123	279	19	12	23	276	9	312	35	21	170	18	33	24

TABLE 18
(Continued)

LNG	SPKR	29	30	31	32	33	34	35	36	37
		FEATURE #								
1	1	241	12	238	156	0	26	216	328	46
1	2	304	14	275	251	0	7	298	337	17
1	3	223	19	194	165	0	17	232	351	17
1	4	197	48	157	169	2	30	160	285	84
1	5	301	19	298	149	1	10	334	353	6
1	6	211	4	194	149	0	0	238	290	0
1	7	273	16	264	232	0	13	229	323	23
1	8	284	18	235	180	2	11	307	345	8
1	9	216	18	195	172	0	10	240	249	11
1	10	234	32	160	158	0	7	238	253	37
2	1	337	17	331	282	36	0	280	350	6
2	2	207	61	184	174	21	30	165	265	43
2	3	250	23	207	197	32	14	247	289	38
2	4	271	17	269	190	10	13	235	330	16
2	5	191	7	175	135	1	3	178	237	5
2	6	222	40	225	187	3	11	204	269	14
2	7	271	40	258	207	3	3	216	305	13
2	8	230	16	177	180	9	5	223	300	12
2	9	216	93	191	200	25	3	189	227	21
2	10	174	45	152	140	9	13	131	232	40
3	1	162	29	168	128	1	26	161	259	35
3	2	217	40	196	150	5	7	195	250	20
3	3	162	25	157	142	1	15	143	208	54
3	4	229	29	220	153	10	17	232	304	15
3	5	173	23	139	114	1	14	161	205	26
3	6	258	37	213	87	8	17	282	309	17
3	7	172	21	160	110	2	28	189	242	42
3	8	172	45	160	154	8	12	152	192	53
3	9	157	26	146	123	4	28	179	199	40
3	10	295	23	208	146	4	8	333	312	22
4	1	220	23	198	144	10	11	186	275	19
4	2	204	13	174	142	3	7	222	248	24
4	3	277	71	247	184	9	37	269	310	108
4	4	210	31	199	216	36	21	142	194	40
4	5	289	32	276	200	15	23	277	348	30
4	6	280	50	233	192	15	22	243	329	62
4	7	305	62	245	181	16	35	313	388	53
4	8	198	48	170	132	7	34	186	245	47
4	9	161	48	152	123	22	30	158	264	49
4	10	296	32	287	245	8	17	290	324	36
5	1	266	35	172	120	5	21	312	357	40
5	2	217	63	151	164	11	14	212	242	29
5	3	290	34	245	230	10	14	256	340	28
5	4	435	63	321	293	4	11	426	413	18
5	5	438	37	368	291	7	11	423	455	38
5	6	294	28	256	235	27	1	287	325	21
5	7	332	41	292	271	5	6	316	390	11
5	8	269	51	231	183	12	11	274	346	30
5	9	252	25	264	199	14	12	194	321	21
5	10	224	23	158	162	2	8	226	270	24

TABLE 19. OCCURRENCE VALUES -- NORMALIZED DATA, TEST SPEAKERS

LNG	SPKR	FEATURE # 1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	10	30	20	89	68	72	2	13	47	29	27	277	109	72
1	2	6	5	16	11	43	69	0	35	43	44	6	244	20	43
1	3	13	16	22	39	44	73	1	16	37	5	22	199	32	45
1	4	7	13	24	29	38	69	0	32	43	22	17	161	21	40
1	5	8	35	31	60	94	95	18	12	19	19	13	174	58	97
1	6	9	11	12	98	36	43	0	9	43	35	9	198	103	41
1	7	5	20	43	63	29	53	5	26	30	22	4	163	28	30
1	8	3	20	21	78	22	28	1	13	56	35	4	212	65	24
1	9	11	9	24	11	29	33	9	34	32	74	1	165	1	31
1	10	10	19	20	8	62	72	2	20	25	26	13	215	3	67
2	1	25	27	44	29	67	79	2	39	78	13	23	297	22	73
2	2	12	9	46	78	26	26	5	8	155	13	20	268	46	26
2	3	30	2	16	80	41	60	13	10	133	19	20	252	90	48
2	4	27	131	76	67	73	32	17	10	124	14	5	236	9	67
2	5	18	57	51	64	23	15	7	9	102	20	3	292	15	23
2	6	43	23	24	32	50	48	37	12	97	13	2	267	13	50
3	1	27	44	61	19	84	111	3	33	132	21	51	209	19	86
3	2	12	23	44	25	85	112	0	37	106	10	25	243	33	91
3	3	12	34	19	23	84	79	0	32	136	55	22	192	24	84
3	4	16	17	49	22	57	95	10	27	93	24	43	126	35	58
3	5	23	60	34	79	95	119	13	15	61	4	39	115	81	107
3	6	16	69	13	35	116	83	0	14	126	37	16	201	31	112
3	7	23	82	34	40	134	85	7	9	133	39	12	189	23	133
3	8	44	26	49	20	70	84	10	42	81	20	19	211	30	70
3	9	11	43	71	32	54	44	2	31	74	22	18	208	15	58
3	10	18	43	47	26	72	61	0	28	86	20	9	148	13	64
4	1	12	15	15	74	62	49	35	6	115	37	7	243	41	63
4	2	8	40	39	68	29	34	6	8	124	24	2	248	29	35
4	3	60	44	62	87	43	21	40	9	154	3	1	217	45	41
4	4	3	11	24	42	4	1	16	7	137	24	0	293	18	3
4	5	16	15	26	23	40	68	14	9	75	7	16	229	19	40
4	6	12	7	40	62	43	42	18	18	94	14	5	313	68	46
4	7	26	6	50	24	13	16	27	8	152	8	0	244	6	16
4	8	22	27	52	46	51	37	31	9	142	31	8	276	23	49
4	9	10	29	86	77	16	9	14	9	101	33	0	260	35	14
4	10	18	53	51	51	51	53	23	23	78	8	3	230	23	52
4	11	22	39	31	89	2	2	9	9	64	2	0	287	11	4
4	12	13	29	42	90	68	54	10	3	34	19	7	306	43	75
4	13	27	24	47	41	41	49	34	7	65	12	2	221	39	44
4	14	28	84	35	90	64	36	11	13	61	3	4	315	25	65
5	1	13	36	21	86	77	68	5	3	98	8	4	271	79	79
5	2	24	80	19	53	56	31	4	13	92	10	10	259	26	53
5	3	11	38	26	62	65	62	8	15	124	17	7	167	25	68
5	4	28	71	18	81	62	25	3	13	148	6	1	193	15	57
5	5	12	36	25	63	67	63	9	13	125	29	6	161	26	68
5	6	35	63	61	53	46	25	6	16	117	23	8	155	26	44
5	7	11	56	18	23	109	84	0	21	85	7	9	214	8	107
5	8	50	25	11	24	66	63	12	12	152	1	9	318	16	69
5	9	43	48	23	64	69	52	1	5	95	30	6	257	35	74
5	10	26	45	38	66	70	83	5	16	158	6	21	222	62	76

TABLE 19
(Continued)

LNG	SPKR	15	16	17	18	19	20	21	22	23	24	25	26	27	28
		FEATURE #													
1	1	91	261	10	4	46	349	2	269	38	11	147	20	43	27
1	2	127	261	7	11	5	241	4	269	30	6	177	30	9	5
1	3	139	251	16	26	15	211	2	226	14	20	195	22	24	19
1	4	180	238	24	35	15	188	3	226	36	23	254	68	15	12
1	5	117	133	17	11	27	226	2	189	28	12	129	16	43	32
1	6	205	253	6	9	42	309	5	248	33	9	253	21	12	18
1	7	115	266	33	40	34	286	15	206	40	48	224	68	22	33
1	8	100	277	22	31	30	358	24	236	57	42	181	35	20	38
1	9	128	197	22	26	5	289	7	191	21	16	211	68	10	10
1	10	153	237	11	13	0	267	12	251	56	7	224	33	31	14
2	1	179	230	26	21	20	330	6	284	42	23	173	24	33	19
2	2	157	277	11	27	60	162	44	274	79	22	214	14	10	30
2	3	119	231	2	2	48	279	11	267	57	4	190	41	6	5
2	4	152	258	83	52	44	246	13	238	26	88	185	14	100	97
2	5	149	254	57	68	26	322	1	280	29	89	179	18	33	69
2	6	198	280	16	31	28	332	41	279	63	45	281	68	23	35
3	1	162	214	47	32	9	290	4	249	48	38	195	50	43	30
3	2	87	215	27	37	16	274	3	237	28	27	135	32	23	22
3	3	100	192	8	6	16	236	7	192	53	5	121	26	46	19
3	4	113	143	24	10	7	322	1	138	38	9	143	30	23	12
3	5	118	156	26	7	57	349	7	134	23	11	135	15	68	50
3	6	205	246	9	8	29	266	7	249	36	13	257	60	79	37
3	7	112	162	21	4	31	389	5	200	24	17	118	13	100	53
3	8	152	241	26	24	13	275	21	243	57	15	224	77	26	8
3	9	63	137	45	25	17	234	2	196	32	41	93	18	42	36
3	10	126	137	41	41	22	372	10	161	30	38	139	40	48	48
4	1	241	261	12	7	35	226	6	283	47	44	267	15	18	37
4	2	133	260	38	48	43	266	13	246	22	53	173	17	24	50
4	3	128	182	28	15	61	230	34	233	77	33	128	23	38	52
4	4	107	289	19	19	27	237	8	321	27	25	124	10	6	24
4	5	216	311	18	15	14	335	9	306	15	12	271	41	13	13
4	6	123	264	12	18	31	225	11	304	38	5	146	10	12	9
4	7	85	164	15	8	12	234	18	216	53	15	98	10	9	15
4	8	148	232	34	26	25	170	24	270	51	31	142	12	32	36
4	9	141	257	67	101	48	216	11	269	18	106	181	15	21	55
4	10	165	252	54	47	47	309	8	248	21	53	226	43	57	58
4	11	249	299	56	109	55	194	15	322	21	93	265	19	15	63
4	12	177	292	23	16	59	224	4	312	30	33	209	17	35	49
4	13	189	184	36	14	21	205	15	217	40	23	181	16	19	28
4	14	303	356	40	31	57	66	9	367	16	61	330	16	75	81
5	1	174	224	16	8	53	275	15	274	60	19	156	8	42	34
5	2	136	261	59	42	48	338	3	259	14	67	151	19	75	83
5	3	69	148	24	18	47	268	8	168	44	41	93	16	51	39
5	4	118	161	18	25	41	362	7	202	39	44	141	26	75	79
5	5	73	141	22	16	41	252	9	158	33	41	95	18	48	43
5	6	86	105	67	40	32	210	12	152	46	45	83	7	58	64
5	7	116	195	18	12	12	214	9	228	42	19	132	31	81	43
5	8	195	236	8	5	15	250	18	338	21	3	175	29	41	14
5	9	195	248	19	9	38	239	14	277	45	26	197	22	54	50
5	10	120	144	30	11	59	253	21	227	36	29	131	27	65	46

TABLE 19
(Continued)

LNG	SPKR	FEATURE # 29	30	31	32	33	34	35	36	37
1	1	195	29	123	101	0	13	228	257	29
1	2	206	25	168	148	0	10	221	266	33
1	3	223	13	175	162	0	6	234	252	13
1	4	258	16	202	222	0	10	238	245	40
1	5	164	14	150	147	0	28	148	191	36
1	6	300	18	230	245	0	8	298	305	7
1	7	234	41	149	166	0	17	260	281	63
1	8	214	61	137	119	2	22	265	282	69
1	9	199	24	148	180	2	15	187	226	41
1	10	248	30	196	192	2	10	239	239	36
2	1	225	24	221	180	23	18	200	301	60
2	2	247	89	178	166	40	17	248	278	65
2	3	219	37	165	182	0	10	184	249	24
2	4	216	51	176	125	3	20	238	257	46
2	5	200	40	181	143	16	26	209	303	52
2	6	280	71	228	247	1	14	252	309	43
3	1	235	41	213	193	7	20	209	233	70
3	2	154	33	116	95	4	38	144	222	66
3	3	150	32	139	109	5	19	152	192	50
3	4	153	24	147	140	8	42	144	159	59
3	5	142	20	149	134	0	25	152	197	39
3	6	270	23	243	229	12	12	259	286	32
3	7	141	24	149	135	6	31	119	205	36
3	8	213	32	192	200	0	23	213	259	66
3	9	109	29	97	71	7	29	102	166	79
3	10	152	42	153	147	7	42	154	180	94
4	1	278	35	265	240	12	6	274	296	14
4	2	200	33	170	105	16	18	229	285	39
4	3	154	71	177	145	35	39	147	199	63
4	4	196	53	141	85	15	5	211	274	18
4	5	305	24	242	241	45	6	312	330	22
4	6	223	52	171	103	11	15	214	293	47
4	7	153	63	123	88	6	22	124	164	54
4	8	175	63	176	136	8	32	170	255	51
4	9	219	68	163	118	19	18	236	294	57
4	10	242	30	219	163	51	23	229	286	58
4	11	290	43	278	225	7	10	283	328	29
4	12	245	42	228	162	14	40	256	335	41
4	13	215	29	229	204	0	18	162	272	33
4	14	367	38	337	286	3	15	361	396	31
5	1	205	48	211	163	13	14	167	275	15
5	2	195	25	168	159	11	10	197	258	24
5	3	100	29	96	84	15	13	95	148	26
5	4	178	30	159	144	10	11	162	196	29
5	5	100	26	101	88	14	14	88	148	27
5	6	94	55	99	94	18	20	63	117	43
5	7	155	36	142	136	1	9	141	194	33
5	8	247	23	238	219	50	11	194	285	13
5	9	262	57	243	232	13	22	229	274	27
5	10	137	53	139	127	34	20	101	181	38

The first stage of FSS evaluates each feature of M by itself. The one feature yielding best accuracy is selected for M_1 and is used in the next stage of FSS. The k-th stage of FSS uses the output $M_{k-1} = \{y_1, y_2, ..., y_{k-1}\}$ of stage k-1. All subsets $\{y_1, ..., y_{k-1}, x\}$, $x \in M-M_k$, are first evaluated selecting y_k *to be the feature allowing the best accuracy.* Then for each of $y_1, ..., y_{k-1}$ substitutions are made one at a time using all features not already included, determining in each case the feature yielding best accuracy. If accuracy is not improved by such substitution the original feature is retained.

The results of this application of the forward search with substitutions are shown in Table 20. The occurence of an asterisk at the intersection of a row and a column indicates the use of the feature corresponding to the column in that subset corresponding to the row. The percent correct classification of the 50 training speakers resulting from use of each subset of features is shown in the last column of the table. For example, the 9-element subset of features $M_9 = \{1,2,5,7,9,10,20,26,34\}$ yielded 94% correct classification.

The six subsets of features yielding the highest training data classification accuracy (94%) were used with the nearest-mean decision rule to classify each of the 50 test speakers. The resultant classification accuracies are:

M_9 - 70%

M_{10} - 74%

M_{11} - 76%

M_{12} - 74%

M_{13} - 72%

M_{14} - 80%.

Based on the use of the 14-element feature subset M_{14}, the 80% classification represents a significant improvement over the accuracy (66%) attained

TABLE 20. RESULTS OF FSS FEATURE SELECTION

	Feature #															
k	1	2	5	7	9	10	11	13	14	20	26	27	30	33	34	% Correct
1						*										46
2					*	*										70
3			*		*	*										74
4			*		*	*									*	82
5	*		*		*	*									*	84
6	*		*		*	*	*								*	86
7	*		*		*	*	*		*						*	86
8	*		*		*	*	*		*	*					*	88
9	*	*	*	*	*	*				*	*				*	94
10	*	*	*	*	*	*	*			*	*				*	94
11	*	*	*	*	*	*	*	*		*	*				*	94
12	*	*	*	*	*	*	*		*	*	*			*	*	94
13		*	*	*	*	*	*	*	*	*	*		*	*	*	94
14	*	*	*	*	*	*	*	*	*	*	*		*	*	*	94
15	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	88

before data normalization. Decision function values and classifications resulting from the use of M_{14} are shown in Table 21 (Training data) and Table 22 (Test data). Comparison of these results with those in Appendix B resulting from the use of original data indicate that the improvement in classification is due mainly to the standardization of the data from L5 speakers. Specifically, following normalization there were no L3 - L5 confusions. The large number of L2 - L4 confusions indicate the need for reference patterns with L2 and L4 language specificity. A final observation is the inclusion of the silence measure in M_{14} indicating its usefulness as a language discriminant.

TABLE 21. DECISION FUNCTION VALUES RESULTING FROM USE OF M_{14}--TRAINING SPEAKERS

DECISION FUNCTION VALUES (DIST. TO LANG. MEAN) (TRAINING DATA)

FEATURES USED: 1 2 5 7 9 10 11 13 14 20 26 30 33 34

LNG	SPKR	L1	L2	L3	L4	L5	DEC.	# CORR.	TOT.
		LANGUAGE							
1	1	2,524	5,368	2,828	4,767	4,746	1		
1	2	2,932	5,474	3,600	5,467	5,147	1		
1	3	2,556	4.235	3,633	3,423	3,934	1		
1	4	3,936	5,169	4,493	4,243	5,320	1		
1	5	2,644	4,369	3,186	4,424	3,963	1		
1	6	4,009	5,842	4,797	5,743	4,902	1		
1	7	2,661	5,350	3,686	4,775	3,905	1		
1	8	2,364	4,753	3,316	4,864	4,446	1		
1	9	2,470	4,959	4,303	4,305	4,664	1		
1	10	2,499	4,477	2,987	4,426	4,404	1		
								10	
2	1	5,373	3,780	5,053	4,853	3,874	2		
2	2	6,244	3,224	5,861	3,698	4,760	2		
2	3	5,340	2,783	5,417	3,350	3,483	2		
2	4	3,980	3,368	4,726	3,464	3,795	2		
2	5	4,425	3,689	5,103	4,398	3,900	2		
2	6	4,622	2,707	4,513	3,914	3,820	2		
2	7	4,208	1,894	4,436	3,328	3,201	2		
2	8	4,152	2,658	4,341	3,765	3,508	2		
2	9	8,667	6,429	8,479	7,144	7,374	2		
2	10	4,130	1,626	3,478	3,127	2,830	2		
								10	
3	1	3,899	5,752	2,399	5,807	5,325	3		
3	2	4,277	4,711	2,893	5,410	4,012	3		
3	3	3,775	5,908	3,185	5,652	5,391	3		
3	4	5,221	6,147	3,683	6,922	6,172	3		
3	5	3,703	5,695	2,698	5,845	4,933	3		
3	6	3,245	2,610	3,329	2,938	3,293	2		
3	7	4,188	6,228	3,595	5,625	5,490	3		
3	8	3,713	4,398	4,166	3,443	4,114	4		
3	9	3,326	4,553	2,079	4,241	4,193	3		
3	10	3,126	3,621	2,691	4,273	3,101	3		
								8	
4	1	3,636	2,823	4,324	2,392	2,481	4		
4	2	3,597	4,246	4,852	3,553	3,927	4		
4	3	5,379	5,432	5,304	3,840	5,007	4		
4	4	6,037	3,414	6,104	3,385	4,379	4		
4	5	3,304	3,222	3,524	2,467	2,736	4		
4	6	4,238	3,435	4,478	1,940	2,544	4		
4	7	5,206	4,269	5,672	2,556	4,282	4		
4	8	4,829	4,076	5,221	2,974	4,085	4		
4	9	4,976	3,257	4,955	1,986	3,186	4		
4	10	4,230	2,686	4,840	2,539	3,870	4		
								10	
5	1	3,783	3,633	3,246	3,331	2,564	5		
5	2	3,902	3,202	4,036	2,462	2,369	5		
5	3	4,053	4,098	3,758	3,914	2,443	5		
5	4	5,673	4,928	5,698	4,765	3,644	5		
5	5	4,689	3,529	5,139	3,641	2,653	5		
5	6	5,547	3,785	5,995	3,674	4,056	4		
5	7	4,626	3,798	4,684	4,098	2,490	5		
5	8	3,960	3,701	3,312	3,758	2,259	5		
5	9	4,646	3,773	4,812	4,387	3,314	5		
5	10	3,127	2,566	3,440	2,883	1,855	5		
								9	47

TABLE 22. DECISION FUNCTION VALUES RESULTING FROM USE OF M_{14}--TEST SPEAKERS

DECISION FUNCTION VALUES (DIST. TO LANG. MEAN) (TEST DATA)

FEATURES USED: 1 2 5 7 9 10 11 13 14 20 26 30 33 34

LNG	SPKR	LANGUAGE L1	L2	L3	L4	L5	DEC.	# CORR.	TOT.
1	1	3.159	5.802	3.303	5.614	5.162	1		
1	2	2.526	4.720	3.761	4.317	4.243	1		
1	3	2.637	4.524	3.759	4.086	3.555	1		
1	4	3.552	5.200	4.442	4.599	4.545	1		
1	5	3.555	5.347	3.781	4.983	4.640	1		
1	6	2.764	5.427	4.219	5.049	5.076	1		
1	7	3.447	5.084	4.498	3.833	4.404	1		
1	8	3.427	5.191	4.273	3.958	4.729	1		
1	9	4.747	6.166	5.615	5.847	6.378	1		
1	10	2.598	4.664	3.109	4.200	3.539	1	10	
2	1	3.674	4.085	2.745	3.792	3.172	3		
2	2	6.964	5.014	6.548	4.919	5.411	4		
2	3	3.776	3.609	3.677	4.072	4.003	2		
2	4	6.049	5.318	5.820	5.379	4.299	5		
2	5	4.081	3.651	4.355	2.363	3.182	4		
2	6	6.399	4.885	6.187	5.075	5.365	2	2	
3	1	5.498	5.761	3.814	6.042	5.180	3		
3	2	4.401	5.449	3.195	4.760	4.315	3		
3	3	3.998	4.885	2.996	5.358	4.454	3		
3	4	4.581	5.403	3.432	4.817	5.362	3		
3	5	4.683	6.256	3.950	6.193	5.416	3		
3	6	5.196	5.990	4.270	6.417	4.935	3		
3	7	6.190	6.953	4.906	7.213	5.981	3		
3	8	4.421	4.883	3.917	4.563	4.261	3		
3	9	3.073	4.197	2.984	3.270	3.131	3		
3	10	4.418	5.498	3.725	4.213	4.292	3	10	
4	1	4.639	2.955	4.663	4.485	4.452	2		
4	2	3.485	2.947	3.882	2.423	2.706	4		
4	3	8.480	5.623	8.028	5.623	6.538	4		
4	4	5.180	3.398	5.729	3.648	4.369	2		
4	5	5.540	4.427	5.383	4.420	4.778	4		
4	6	3.664	2.764	4.068	2.659	3.158	4		
4	7	6.041	3.489	6.091	3.451	4.493	4		
4	8	5.595	3.470	5.261	3.813	4.509	2		
4	9	4.715	3.403	5.236	2.835	3.986	4		
4	10	6.488	5.033	6.319	4.669	5.294	4		
4	11	4.644	3.871	5.692	2.770	3.429	4		
4	12	3.913	4.792	3.956	3.451	3.984	4		
4	13	4.399	3.270	5.071	3.365	4.000	2		
4	14	4.615	4.417	5.365	4.109	2.946	5	9	
5	1	3.608	4.178	3.683	4.015	2.986	5		
5	2	3.538	3.916	3.589	3.686	2.366	5		
5	3	3.283	2.804	2.937	3.119	1.902	5		
5	4	4.506	4.001	4.145	4.011	2.493	5		
5	5	3.115	2.779	2.846	3.296	2.325	5		
5	6	4.562	3.201	4.422	2.959	2.518	5		
5	7	4.702	5.534	4.247	5.582	3.653	5		
5	8	7.081	4.748	6.409	5.343	4.932	2		
5	9	4.231	4.055	3.798	3.736	2.840	5		
5	10	5.620	4.252	4.583	4.426	3.803	5	9	40

SECTION V
EPILOGUE

A. Summary and Conclusions

A study of consonant-vowel-consonant hyperphones in English indicated that such 3-phoneme sequences do not occur often enough to provide operational capability. Consideration was given to the determination of recurrences of a broad sound class rather than a specific phoneme-like sound. Six classes of sound transitions were considered. It was found that the continuous nature of the data precluded reliable detection of component transitions using 50 ms scanning patterns. The use of single steady-state reference sounds was then studied. An automatic speaker adaptation technique provided modification of standard references to allow for small speaker-to-speaker spectral variations. This technique allowed stable detection of English nasals and sibilants in continuous speech. Departing from completely automatic reference selection techniques used in previous studies, interactive reference extraction procedures were applied. Preliminary processing of 36 references indicated that they possessed some language specificity. A five-language classification experiment using 13 of these references, along with a silence measure, yielded 17 classification errors of the 50 test speakers classified (66% correct classification).

Standardization of the long-term average spectrum for each speaker in the data base and subsequent data reprocessing using the same 36 references provided much improved performance. Use of a second set of 13 reference sounds along with the same silence measure provided classification of the 50 test speakers with only 10 errors, providing 80% correct classification. This improved performance provides several implications. Long-term average spectra vary much less with language than with speaker and/or recording conditions. Several single steady-state reference sounds have been determined which provide good language classification. The number of references required is small enough to provide real-time operation of a classifier. And, an appropriate voicing measure will aid in discriminating languages.

B. Recommendations

The collection of language-specific reference sounds should be expanded in several directions. More sounds useful for discriminating languages L2 and L4 are needed. Use of pairs of sounds rather than single sounds is potentially valuable. Multiple representations of established references would help alleviate the problem of speaker dependence.

In view of the coloration of previously obtained results[2] by the non-uniform recording conditions for collection of the data base, the reference selection techniques and resultant reference sounds should be reconsidered. The poor results from using those techniques may have been due more to the spectral coloration than to the nature of the resultant references.

Consideration should be given to development of a speaker adaptation procedure. Use of speaker adaptation would provide better detection of reference sounds and keep smaller the number of required references. Specific areas for research include developing a rule for cessation of adaptation, and determining a method for preventing transmutation of the adapted sound away from the target sound.

Other measures of language discrimination should be considered. For language L1 and L3, pitch is phonemic. For languages L2, L4, and L5 pitch conveys non-phonemic information. The use of pitch as a language discriminant could provide a much more robust language classification algorithm. The observation that a silence-to-speech ratio was useful for closed-set language identification warrants further investigation into use of this measure and similar voicing interval measures for language classification.

APPENDIX A

CANDIDATE REFERENCE SCANNING PATTERNS

LEGEND

```
SYMBOL     , " + = 0 B $

LEVEL    0 1 2 3 4 5 6 7
```

```
(+B+  0+  )(+ B)
(+0=, 0+" )(0 D)
(=+=0=+=B=)(B=+)

REF  1  ( 105  25)
```

```
(  0B0  "+)(  B)
( ,BB= ,"")(  B)
(==0=+====)(==+)

REF  2  ( 101  35)
```

```
( 0$"+ ,",)(" B)
( 0$"= ,",)(" B)
(=========)(==+)

REF  3  ( 108  25)
```

```
(  +B,",",=)(  0)
( =B ",",+)(  0)
(==0=====+)(===)

REF  4  ( 106 130)
```

```
(  ,0B  +0)(  B)
(  ,0B  +=)(  B)
(==0======)(===)

REF  5  ( 110  40)
```

```
("  =B+ "0)(  B)
("  =0" +0)(  0)
(=========)(===)

REF  6  ( 112  34)
```

```
(+0""B ,,")(" 0)
("=""B  "")(, B)
(++=====00)(+=0)

REF  7  ( 113 [illegible])
```

```
(,B0+, "" )(B =)
( 00=, "" )(B =)
(++=0=====)(+==)

REF  8  (  92  43)
```

```
(  "=0" ,$)($$=)
(  "0B+  $)($$B)
(===000=+=)(==B)

REF  9  ( 160   1)
```

```
("",",=+""+)("0 )
(,,,+=",",=)($$+)
("+=0=++00)($B$)

REF 10  ( 115  17)
```

```
(+  B,= "+)(  B)
(+  =B, "+)(  B)
(==="$,===)(===)

REF 11  ( 116   9)
```

```
(B+   ,+0" )($"0)
(B=   "=0, )($ B)
(=0===0=+=)(="0)

REF 12  ( 120 295)
```

```
(  "B = "=)(  B)
(,  B = "=)(  0)
(==+======)(==+)

REF 13  ( 122  27)
```

```
(  ,0B  +0)(  B)
(  ,0B, +0)(  B)
(=========)(==0)

REF 14  ( 127  22)
```

```
(0$   0=  )($ 0)
(0$,  0=  )($ $)
(==0======)(==B)

REF 15  ( 130 105)
```

```
(B0  +++, )($ $)
($0  +=+, )($ $)
(=====0===)(===)

REF 16  ( 131 150)
```

```
( +$0+ ,,,)(, B)
( +$== ,"")(, $)
(===+0====)(==0)

REF 17  ( 132  37)
```

```
( +$0 ,,,")(" $)
( +$=, ,"")(" 0)
(===+0+=0=)(==")

REF 18  ( 133  51)
```

```
(  +B"+ ,+)(  $)
( ,=$ ","")(  $)
(=0B0"=00+)(===)

REF 19  ( 134  62)
```

```
(,BB+ =+, )(+ $)
(+B0+ =+  )(= $)
(B==0=====)(B==)

REF 21  ( 136  26)
```

```
(B0  ,+0" )($ B)
($=  ,=0" )($ B)
(=+===0===)(===)

REF 22  ( 138 129)
```

```
( 0$" =+" )(= $)
(,B$, =+" )(B $)
(00=======)(B==)

REF 23  ( 139  27)
```

```
( "$0+" +")(  $)
( "$0", "")(  $)
(=========)(===)

REF 24  ( 140  43)
```

```
(0$,,,++, )($ $)
(0$,,,++, )($ $)
(=========)(==0)

REF 25  ( 142 160)
```

```
(=$=","", )($ $)
(=B==,,", )($ B)
(===B=+===)(===)

REF 26  ( 143   9)
```

```
(  +BB  "=)(  B)
(  0B0  "+)(  $)
(==B0+====)(==0)

REF 27  ( 141  21)
```

```
( ,0B=  "=)(  B)
( ,BB"" "+)(  B)
(==0="B===)(===)

REF 28  ( 161  34)
```

```
(B$, ,==" )($ $)
(B$, ,+=, )($ 0)
(==0=0====)(==")

REF 29  ( 145 146)
```

```
(,B$, =+, )(+ B)
(,0B" =,, )(+ B)
(=++0==+0=)(===)

REF 30  ( 146  27)
```

```
(0B   0=, )($ =)
(0$,  0=  )($ B)
(=00====+=)(==B)

REF 31  ( 147 139)
```

```
(=$"  ==  )(B B)
(=$"  ==  )(B B)
(======+==)(==+)

REF 32  ( 148 103)
```

```
( ""==  = )($$0)
( ""++ "0 )(B$B)
(===++=B0=)(+=0)

REF 33  ( 150  12)
```

```
( $$ +"", )(0 $)
( 0$"0 ,,,)(" B)
(="=BB++=0)("=+)

REF 34  ( 151  15)
```

```
(B$   +=+, )($ $)
(B$   +++, )($ 0)
(=====+===)(==")

REF 35  ( 153 229)
```

```
(B0   "==, )($ +)
(BB   "+=, )($ $)
(=0=====+=)(==$)

REF 36  ( 155 202)
```

```
(,BB"",,, )(0 B)
( B$"" ," )(= B)
(==0====0=)(+==)

REF 37  ( 158  38)
```

APPENDIX B

DECISION FUNCTION VALUES

DECISION FUNCTION VALUES (DIST. TO LANG. MEAN) (TRAINING DATA)

FEATURES USED: 6 9 10 11 19 24

		LANGUAGE							
LNG	SPKR	L1	L2	L3	L4	L5	DEC.	# CORR.	TOT.
1	1	1.533	3.160	2.873	4.254	4.193	1		
1	2	1.574	3.256	3.250	3.549	3.238	1		
1	3	1.594	2.145	2.551	2.053	2.900	1		
1	4	1.538	2.166	2.098	2.520	2.909	1		
1	5	1.153	2.195	2.134	2.037	2.410	1		
1	6	2.524	4.357	4.228	5.391	5.108	1		
1	7	1.475	2.583	2.631	2.727	3.353	1		
1	8	2.075	3.903	3.699	4.944	4.707	1		
1	9	1.903	3.426	3.437	2.927	2.713	1		
1	10	2.075	3.012	3.205	2.113	3.316	1		
								10	
2	1	2.667	1.192	1.858	3.337	3.136	2		
2	2	3.312	1.053	2.015	2.290	2.481	2		
2	3	2.644	1.622	1.967	1.900	1.782	2		
2	4	1.680	1.530	2.261	2.885	2.850	2		
2	5	2.329	1.150	1.516	2.726	2.749	2		
2	6	3.454	1.720	2.329	4.033	3.621	2		
2	7	3.492	1.313	2.125	2.570	2.945	2		
2	8	3.649	3.304	3.422	4.535	3.575	2		
2	9	3.593	1.715	1.725	2.530	2.871	2		
2	10	2.544	0.654	1.300	2.432	2.594	2		
								10	
3	1	3.461	2.905	1.835	4.550	4.145	3		
3	2	1.784	1.582	1.161	2.665	2.101	3		
3	3	2.995	2.256	1.714	2.520	3.155	3		
3	4	3.223	2.515	2.116	4.609	3.842	3		
3	5	2.366	2.051	0.983	3.394	3.003	3		
3	6	2.617	1.407	1.747	1.568	2.129	2		
3	7	3.300	1.489	1.022	3.143	3.038	3		
3	8	3.538	1.631	1.292	3.261	2.872	3		
3	9	3.058	2.298	1.608	3.260	2.159	3		
3	10	2.737	1.681	1.666	2.141	2.828	3		
								9	
4	1	3.310	3.455	3.629	1.211	2.615	4		
4	2	2.924	2.294	2.781	1.485	2.832	4		
4	3	3.159	2.128	2.449	1.381	2.562	4		
4	4	3.371	3.103	3.390	0.731	1.675	4		
4	5	1.778	1.962	2.425	1.336	2.044	4		
4	6	3.398	2.976	3.201	1.147	2.183	4		
4	7	3.279	3.578	3.787	1.530	2.189	4		
4	8	3.744	2.116	2.391	1.639	1.705	4		
4	9	3.708	3.350	3.638	1.372	1.389	4		
4	10	3.267	2.890	3.203	1.040	1.438	4		
								10	
5	1	2.459	1.674	2.426	1.753	1.346	5		
5	2	2.447	1.803	2.418	1.388	1.257	5		
5	3	1.993	2.519	2.775	1.973	2.054	4		
5	4	6.724	6.576	6.549	5.191	4.435	5		
5	5	4.214	3.501	3.793	2.210	1.344	5		
5	6	3.229	2.809	3.161	1.643	0.812	5		
5	7	3.525	3.413	3.788	2.763	2.120	5		
5	8	3.067	3.587	3.599	2.782	1.804	5		
5	9	3.688	2.351	2.362	2.460	1.332	5		
5	10	6.212	4.354	3.793	5.568	5.378	3		
								8	47

DECISION FUNCTION VALUES (DIST. TO LANG. MEAN) (TEST DATA)

FEATURES USED: 6 9 10 11 19 24

		LANGUAGE							
LNG	SPKR	L1	L2	L3	L4	L5	DEC.	# CORR.	TOT.
1	1	3.138	4.929	4.962	5.716	5.473	1		
1	2	2.643	2.705	2.352	2.211	3.200	4		
1	3	1.606	2.459	2.488	1.576	2.404	4		
1	4	2.327	2.491	2.308	1.733	2.570	4		
1	5	2.064	3.991	3.808	4.724	4.469	1		
1	6	2.528	3.132	3.204	1.557	2.089	4		
1	7	3.193	3.513	3.476	1.492	2.499	4		
1	8	2.944	2.716	2.610	1.057	2.181	4		
1	9	2.710	2.779	2.356	1.691	2.507	4		
1	10	2.073	2.685	2.743	2.239	3.135	1		
								3	
2	1	1.666	1.704	1.604	3.099	2.619	3		
2	2	3.398	2.439	2.708	1.933	1.097	5		
2	3	2.446	3.070	3.232	2.795	2.341	5		
2	4	5.105	4.559	4.583	3.159	2.726	5		
2	5	3.543	2.674	2.690	1.640	1.436	5		
2	6	3.133	1.882	2.060	1.395	2.035	4		
								0	
3	1	3.460	1.554	1.151	3.466	3.292	3		
3	2	3.471	1.315	1.289	3.156	3.049	3		
3	3	3.409	1.482	1.230	3.572	3.292	3		
3	4	2.101	1.420	0.997	3.274	3.150	3		
3	5	3.006	1.634	1.108	3.474	3.054	3		
3	6	4.139	2.379	1.993	4.401	3.902	3		
3	7	2.008	1.754	1.603	3.784	3.198	3		
3	8	3.765	1.364	1.963	2.801	2.640	2		
3	9	3.105	2.195	2.592	1.301	2.583	4		
3	10	2.943	1.781	1.624	1.494	1.780	4		
								7	
4	1	1.408	1.642	1.411	2.852	2.975	1		
4	2	4.065	3.397	3.633	1.795	1.830	4		
4	3	3.539	2.620	3.183	1.462	1.219	5		
4	4	3.396	2.216	2.772	1.022	2.140	4		
4	5	2.547	2.271	2.711	1.668	2.939	4		
4	6	1.900	2.133	2.348	1.426	2.245	4		
4	7	3.186	2.362	2.889	1.388	2.740	4		
4	8	3.105	2.473	2.250	1.293	1.643	4		
4	9	4.072	3.238	3.437	1.578	1.383	5		
4	10	3.427	2.972	3.135	1.271	2.155	4		
4	11	3.273	2.692	3.106	0.424	1.967	4		
4	12	1.672	3.011	2.984	2.224	2.170	1		
4	13	2.628	1.984	2.385	1.340	2.487	4		
4	14	5.055	5.393	5.484	3.478	3.388	5		
								9	
5	1	3.758	1.645	1.603	3.406	3.155	3		
5	2	6.368	5.897	6.016	4.824	4.033	5		
5	3	3.083	1.794	2.004	1.452	1.089	5		
5	4	4.768	3.202	2.578	3.526	3.230	3		
5	5	3.413	1.736	1.553	2.145	1.823	3		
5	6	3.771	2.105	1.832	2.447	2.676	3		
5	7	2.559	1.254	1.326	2.421	2.199	2		
5	8	3.946	1.674	1.850	3.140	3.175	2		
5	9	3.715	2.736	1.829	3.134	3.178	3		
5	10	4.119	1.761	2.040	3.245	3.190	2		
								2	21

DECISION FUNCTION VALUES (DIST. TO LANG. MEAN) (TRAINING DATA)

FEATURES USED: 2 7 9 10 20 27 34

		LANGUAGE							
LNG	SPKR	L1	L2	L3	L4	L5	DEC.	# CORR.	TOT.
1	1	1.402	2.326	2.897	2.497	3.237	1		
1	2	1.702	3.585	2.422	2.186	1.753	1		
1	3	1.529	2.965	2.762	1.610	2.972	1		
1	4	1.902	2.731	2.211	1.077	2.454	4		
1	5	0.840	3.165	2.456	2.879	2.776	1		
1	6	1.785	4.181	2.883	3.393	3.687	1		
1	7	1.278	3.058	3.021	2.426	3.139	1		
1	8	1.241	2.592	3.174	2.891	3.328	1		
1	9	3.016	4.630	4.145	4.359	3.465	1		
1	10	2.166	3.796	4.227	3.862	4.630	1		
								9	
2	1	3.178	1.116	3.125	3.277	3.080	2		
2	2	4.046	1.753	4.001	3.228	3.749	2		
2	3	3.535	2.490	3.824	3.627	2.686	2		
2	4	3.807	2.531	4.882	4.701	4.720	2		
2	5	1.917	1.828	1.841	2.635	1.923	2		
2	6	4.428	1.688	4.139	4.346	4.157	2		
2	7	3.192	0.844	3.025	3.159	3.177	2		
2	8	3.115	1.661	2.719	3.786	2.795	2		
2	9	3.260	2.360	2.203	2.413	3.274	3		
2	10	2.779	0.752	3.009	2.913	3.303	2		
								9	
3	1	3.736	3.965	1.527	3.875	3.492	3		
3	2	2.728	3.725	1.830	3.277	2.183	3		
3	3	2.399	3.498	1.177	2.449	2.896	3		
3	4	3.217	2.613	1.098	3.194	2.609	3		
3	5	3.058	4.119	2.252	4.255	3.759	3		
3	6	1.998	2.373	1.569	1.984	2.251	3		
3	7	2.972	2.668	1.130	3.191	2.774	3		
3	8	3.299	3.104	1.531	3.599	3.181	3		
3	9	5.156	5.307	3.800	4.188	3.829	3		
3	10	2.621	2.572	2.539	2.690	2.910	3		
								10	
4	1	1.738	3.653	3.011	1.484	2.838	4		
4	2	1.839	2.728	2.634	1.748	2.798	4		
4	3	2.320	3.108	2.220	1.677	3.177	4		
4	4	1.915	3.041	2.640	1.025	1.884	4		
4	5	2.525	2.993	3.151	0.967	2.300	4		
4	6	2.693	3.447	2.928	0.695	2.351	4		
4	7	4.642	5.117	5.086	2.786	4.803	4		
4	8	3.744	2.874	2.916	1.955	2.691	4		
4	9	2.392	3.136	2.786	1.464	1.587	4		
4	10	2.332	2.844	2.800	0.707	2.458	4		
								10	
5	1	2.378	3.380	2.236	2.531	2.130	5		
5	2	2.625	2.288	2.238	1.497	1.262	5		
5	3	2.869	4.028	3.200	2.264	1.702	5		
5	4	6.238	6.774	6.062	5.960	4.506	5		
5	5	3.839	3.109	3.790	2.919	2.003	5		
5	6	2.287	2.891	2.582	2.314	1.168	5		
5	7	2.647	2.719	3.373	2.375	2.226	5		
5	8	2.809	3.904	3.154	2.671	1.498	5		
5	9	4.315	3.790	3.320	4.359	2.153	5		
5	10	5.689	4.559	3.918	5.768	5.217	3		
								9	47

DECISION FUNCTION VALUES (DIST. TO LANG. MEAN) (TEST DATA)

FEATURES USED: 2 7 9 10 20 27 34

LNG	SPKR	L1	L2	L3	L4	L5	DEC.	# CORR.	TOT.
		LANGUAGE							
1	1	0.918	3.476	2.987	2.595	3.215	1		
1	2	2.172	3.284	2.465	2.186	3.258	1		
1	3	1.637	3.180	2.815	1.984	2.484	1		
1	4	2.295	3.183	2.850	1.294	2.911	4		
1	5	1.796	3.107	3.481	2.956	2.890	1		
1	6	1.239	3.472	2.681	2.140	2.894	1		
1	7	2.218	3.767	3.162	1.174	3.084	4		
1	8	1.685	3.471	2.077	1.795	2.990	1		
1	9	1.373	3.280	1.998	2.263	2.951	1		
1	10	1.158	3.076	2.869	1.746	3.079	1		
								8	
2	1	1.740	2.594	1.273	2.238	1.564	3		
2	2	3.568	2.810	3.094	1.839	3.018	4		
2	3	1.788	2.529	2.807	1.441	3.034	4		
2	4	4.687	5.007	4.125	3.326	2.925	5		
2	5	1.671	2.298	1.299	1.824	2.051	3		
2	6	2.070	1.912	2.327	2.400	3.031	2		
								1	
3	1	3.398	2.738	1.290	3.156	2.665	3		
3	2	3.489	2.632	1.809	2.727	2.902	3		
3	3	3.555	2.771	1.646	3.013	2.544	3		
3	4	2.148	2.142	1.210	2.250	2.589	3		
3	5	3.649	3.971	2.127	4.189	2.803	3		
3	6	4.631	4.054	2.637	4.507	2.989	3		
3	7	3.958	4.355	2.761	4.115	2.961	3		
3	8	4.054	3.409	2.511	2.729	2.708	3		
3	9	2.080	2.830	2.137	0.942	2.356	4		
3	10	2.364	3.278	0.973	2.159	2.237	3		
								9	
4	1	2.548	1.707	3.369	3.456	3.667	2		
4	2	2.695	3.330	2.578	1.029	2.256	4		
4	3	5.576	5.144	5.715	4.891	4.926	4		
4	4	2.223	2.638	2.499	2.325	2.930	1		
4	5	1.508	3.173	2.765	2.457	3.461	1		
4	6	1.697	2.465	2.684	1.293	2.631	4		
4	7	2.253	2.322	3.004	1.335	3.045	4		
4	8	5.293	4.792	5.221	4.138	4.676	4		
4	9	3.348	3.550	3.064	1.532	3.448	4		
4	10	3.208	4.059	3.254	1.605	3.301	4		
4	11	2.270	2.855	3.006	1.924	2.937	4		
4	12	1.712	2.939	3.078	1.867	2.491	1		
4	13	3.570	3.330	3.776	1.858	3.739	4		
4	14	3.461	4.268	4.412	3.282	2.904	5		
								9	
5	1	3.497	2.585	1.773	3.437	2.544	3		
5	2	2.592	3.377	1.882	2.804	1.415	5		
5	3	2.448	2.712	1.519	2.233	1.093	5		
5	4	4.213	4.116	1.918	4.310	3.148	3		
5	5	2.940	2.708	1.129	2.635	1.594	3		
5	6	3.430	2.881	1.833	2.498	2.068	3		
5	7	3.788	3.877	3.027	3.714	1.901	5		
5	8	3.501	2.699	1.830	3.400	3.013	3		
5	9	3.507	3.493	1.797	3.223	2.946	3		
5	10	3.626	2.741	2.143	3.016	3.145	3		
								3	30

DECISION FUNCTION VALUES (DIST. TO LANG. MEAN) (TRAINING DATA)

FEATURES USED: 6 7 9 10 11 20 21 24 27 28 33 34

LNG	SPKR	L1	L2	L3	L4	L5	DEC.	# CORR.	TOT.
		LANGUAGE							
1	1	1.916	3.419	3.810	5.274	4.981	1		
1	2	2.035	4.188	3.452	4.543	3.674	1		
1	3	2.018	3.150	3.117	3.183	3.729	1		
1	4	2.815	3.333	3.337	3.074	3.411	1		
1	5	1.493	3.256	2.556	3.929	3.582	1		
1	6	3.085	5.611	4.646	6.648	6.145	1		
1	7	1.815	3.333	3.351	4.122	4.120	1		
1	8	2.724	4.145	4.333	5.777	5.551	1		
1	9	2.732	4.546	4.210	5.300	4.248	1		
1	10	2.852	3.979	4.535	4.645	5.132	1		
								10	
2	1	3.647	1.796	3.203	4.419	4.093	2		
2	2	4.582	2.011	4.243	3.689	4.099	2		
2	3	4.178	2.787	4.158	4.165	3.011	2		
2	4	3.904	2.839	4.991	5.669	5.478	2		
2	5	2.488	2.266	2.385	4.208	3.449	2		
2	6	4.449	2.235	4.482	5.564	4.995	2		
2	7	3.913	1.656	3.612	3.944	3.843	2		
2	8	3.714	2.965	3.569	5.391	4.578	2		
2	9	5.334	4.028	4.317	3.433	4.277	4		
2	10	3.343	1.210	3.004	3.678	3.978	2		
								9	
3	1	5.317	5.287	3.304	5.546	5.604	3		
3	2	2.973	3.834	2.191	4.125	2.814	3		
3	3	3.906	4.030	2.097	3.184	4.108	3		
3	4	3.892	3.656	2.364	5.191	4.640	3		
3	5	3.183	4.274	2.516	5.313	4.564	3		
3	6	3.010	2.643	2.155	2.564	2.921	3		
3	7	3.519	2.855	1.512	4.093	3.780	3		
3	8	3.781	3.411	2.007	4.495	3.671	3		
3	9	5.261	5.368	3.997	4.848	4.080	3		
3	10	3.443	2.858	2.847	3.498	3.636	3		
								10	
4	1	4.353	4.684	4.428	2.029	3.252	4		
4	2	4.622	4.240	3.858	3.064	4.585	4		
4	3	4.091	3.806	2.991	2.165	3.965	4		
4	4	5.547	5.135	4.863	2.433	3.653	4		
4	5	3.350	3.199	3.261	1.977	3.110	4		
4	6	4.646	4.384	3.994	1.101	2.757	4		
4	7	6.147	5.868	5.835	2.946	5.238	4		
4	8	4.928	3.633	3.959	2.614	3.092	4		
4	9	4.645	4.256	4.071	1.652	2.247	4		
4	10	4.135	3.789	4.110	2.020	2.869	4		
								10	
5	1	2.966	3.532	2.653	3.245	2.397	5		
5	2	3.085	2.605	2.720	1.918	1.968	4		
5	3	3.382	4.080	3.385	3.078	2.369	5		
5	4	9.188	9.155	8.677	7.660	6.249	5		
5	5	5.686	4.710	5.377	3.928	2.599	5		
5	6	5.007	4.653	4.690	3.371	2.603	5		
5	7	3.444	2.944	3.742	3.190	2.521	5		
5	8	4.098	4.631	3.930	3.584	2.103	5		
5	9	5.131	4.349	4.020	4.968	2.696	5		
5	10	6.481	4.909	4.446	6.295	5.828	3		
								8	47

DECISION FUNCTION VALUES (DIST. TO LANG. MEAN) (TEST DATA)

FEATURES USED: 6 7 9 10 11 20 21 24 27 28 33 34

LNG	SPKR	L1	L2	L3	L4	L5	DEC.	# CORR.	TOT.
		LANGUAGE							
1	1	3.208	5.512	5.221	6.596	6.138	1		
1	2	3.309	3.680	3.147	3.295	3.924	3		
1	3	2.475	3.451	3.211	3.178	2.916	1		
1	4	3.218	3.573	3.325	2.860	3.405	4		
1	5	2.745	4.411	4.615	5.907	5.004	1		
1	6	2.422	3.717	3.172	2.954	3.166	1		
1	7	4.949	5.250	5.155	2.996	3.780	4		
1	8	3.765	4.244	3.536	2.537	3.228	4		
1	9	3.039	3.752	2.927	2.865	2.999	4		
1	10	2.450	3.351	3.399	3.009	3.604	1		
								5	
2	1	2.222	3.062	1.814	3.966	2.968	3		
2	2	7.464	6.372	6.299	4.627	5.876	4		
2	3	2.761	3.192	3.747	3.240	3.600	1		
2	4	7.088	6.818	6.216	4.666	4.049	5		
2	5	4.004	3.754	2.980	2.775	2.299	5		
2	6	3.677	2.668	2.901	2.478	3.228	4		
								0	
3	1	3.847	2.916	1.357	3.933	3.534	3		
3	2	3.902	2.765	1.916	3.558	3.497	3		
3	3	4.566	3.547	2.235	4.109	4.071	3		
3	4	2.991	2.838	1.672	3.754	3.933	3		
3	5	4.718	4.667	2.798	5.230	4.312	3		
3	6	5.330	4.547	3.086	5.498	4.489	3		
3	7	4.169	4.708	3.222	5.567	4.220	3		
3	8	4.600	3.658	3.020	3.586	3.277	3		
3	9	3.612	3.354	2.992	1.890	3.097	4		
3	10	4.161	4.062	2.220	2.687	3.219	3		
								9	
4	1	2.669	1.946	3.412	4.491	4.410	2		
4	2	5.410	5.020	4.368	2.197	3.138	4		
4	3	8.058	7.037	7.497	5.845	6.340	4		
4	4	4.081	3.440	3.476	2.370	3.183	4		
4	5	3.207	3.610	3.110	3.010	4.224	4		
4	6	3.608	3.338	3.254	2.326	3.812	4		
4	7	3.869	3.143	3.965	2.234	3.451	4		
4	8	6.269	5.314	5.774	4.230	4.748	4		
4	9	7.197	6.512	6.381	4.032	5.255	4		
4	10	6.846	6.541	5.690	3.926	5.535	4		
4	11	4.062	3.661	3.894	2.114	2.898	4		
4	12	3.715	3.929	3.716	3.070	3.792	4		
4	13	4.409	3.732	4.299	2.734	4.012	4		
4	14	5.872	5.841	6.180	4.395	3.661	5		
								12	
5	1	4.040	2.923	2.311	4.455	3.530	3		
5	2	5.639	5.665	4.880	4.420	3.151	5		
5	3	4.275	3.733	2.657	2.663	1.999	5		
5	4	5.760	5.094	3.341	4.765	3.925	3		
5	5	4.426	3.581	2.291	2.964	2.221	5		
5	6	4.460	3.436	2.726	3.095	2.719	5		
5	7	4.298	4.134	3.341	4.555	2.782	5		
5	8	5.693	4.475	3.573	4.516	4.994	3		
5	9	5.626	5.090	3.749	4.227	4.889	3		
5	10	4.598	3.030	2.475	3.314	3.593	3		
								5	31

DECISION FUNCTION VALUES (DIST. TO LANG. MEAN) (TRAINING DATA)

FEATURES USED: 6 7 8 9 10 13 19 20 22 24 27 28 33 34

		LANGUAGE							
LNG	SPKR	L1	L2	L3	L4	L5	DEC.	# CORR.	TOT.
1	1	2.364	3.174	3.817	4.977	5.063	1		
1	2	2.395	4.403	3.776	4.052	3.196	1		
1	3	2.530	3.375	3.826	3.206	4.119	1		
1	4	2.622	3.339	2.944	2.761	3.588	1		
1	5	1.210	3.379	3.104	4.062	3.778	1		
1	6	2.723	5.142	4.093	6.053	5.724	1		
1	7	2.013	3.374	3.578	4.532	4.772	1		
1	8	2.670	3.794	4.100	5.471	5.476	1		
1	9	4.042	5.865	5.643	6.295	5.212	1		
1	10	2.934	4.027	4.891	4.886	5.538	1		
								10	
2	1	4.067	2.296	4.119	4.711	4.696	2		
2	2	4.388	1.907	4.358	4.162	4.647	2		
2	3	3.945	2.671	4.298	4.372	3.431	2		
2	4	4.082	3.256	5.414	5.805	5.749	2		
2	5	2.733	2.591	2.742	4.934	4.438	2		
2	6	4.593	1.982	4.456	5.564	5.256	2		
2	7	4.068	1.980	4.150	4.404	4.578	2		
2	8	4.337	3.751	4.201	5.278	4.300	2		
2	9	4.137	2.758	2.977	3.307	4.359	2		
2	10	3.527	1.399	3.217	3.843	4.430	2		
								10	
3	1	5.539	5.450	3.278	5.628	5.920	3		
3	2	2.997	3.902	2.252	4.003	2.847	3		
3	3	4.671	4.872	2.663	4.018	5.122	3		
3	4	4.049	3.910	2.668	5.140	4.731	3		
3	5	3.499	4.378	2.769	5.620	5.114	3		
3	6	3.025	2.678	2.559	2.805	3.346	3		
3	7	3.662	3.095	1.845	4.745	4.633	3		
3	8	4.089	3.781	2.135	4.752	4.320	3		
3	9	6.616	6.997	5.309	5.792	5.319	3		
3	10	3.449	2.832	3.223	3.937	4.315	2		
								9	
4	1	4.993	5.534	5.102	2.770	4.128	4		
4	2	4.595	4.362	3.789	3.410	5.057	4		
4	3	4.222	4.133	3.443	2.111	4.254	4		
4	4	4.991	4.936	4.571	2.129	3.438	4		
4	5	3.487	3.441	3.800	1.756	3.208	4		
4	6	4.914	4.827	4.442	1.667	3.427	4		
4	7	6.644	6.840	6.549	3.666	5.618	4		
4	8	4.431	3.579	3.853	3.069	3.381	4		
4	9	4.683	4.609	4.304	2.079	2.310	4		
4	10	3.682	4.055	4.160	2.528	2.928	4		
								10	
5	1	3.399	3.918	3.406	3.287	2.550	5		
5	2	3.021	2.810	3.043	1.960	2.075	4		
5	3	4.042	4.870	4.253	3.168	2.674	5		
5	4	9.345	9.534	9.025	7.862	6.306	5		
5	5	5.949	5.283	5.977	4.255	2.830	5		
5	6	4.939	5.077	4.432	3.232	2.505	5		
5	7	4.425	4.589	5.183	4.151	3.159	5		
5	8	4.744	5.376	4.810	3.791	2.243	5		
5	9	5.277	5.036	4.679	5.816	3.716	5		
5	10	6.555	5.044	4.583	6.820	6.518	3		
								8	47

DECISION FUNCTION VALUES (DIST. TO LANG. MEAN) (TEST DATA)

FEATURES USED: 6 7 8 9 10 13 19 20 22 24 27 28 33 34

LNG	SPKR	L1	L2	L3	L4	L5	DEC.	# CORR.	TOT.
		LANGUAGE							
1	1	1.916	4.507	4.220	5.195	4.896	1		
1	2	3.089	3.591	3.297	3.694	4.497	1		
1	3	2.360	3.451	3.302	3.563	3.527	1		
1	4	2.960	3.468	3.300	3.134	3.866	1		
1	5	2.779	4.327	4.644	5.776	5.000	1		
1	6	3.503	5.154	4.976	4.474	4.367	1		
1	7	3.641	4.465	4.117	2.767	3.670	4		
1	8	2.624	3.928	3.254	2.948	3.611	1		
1	9	2.768	3.721	2.811	3.355	3.649	1		
1	10	2.233	3.323	3.534	3.685	4.405	1		
								9	
2	1	2.512	3.130	2.076	4.215	3.498	3		
2	2	6.919	6.051	5.848	4.477	5.587	4		
2	3	3.488	4.508	4.743	4.471	4.402	1		
2	4	7.039	6.917	6.246	4.784	4.173	5		
2	5	3.740	3.689	3.261	3.119	2.705	5		
2	6	3.460	2.591	3.214	3.064	3.874	2		
								1	
3	1	4.128	3.058	1.878	4.270	4.235	3		
3	2	4.048	2.957	2.170	4.166	4.354	3		
3	3	4.663	3.858	2.560	4.761	4.942	3		
3	4	3.558	3.138	1.946	4.188	4.674	3		
3	5	5.411	5.935	4.445	6.739	6.007	3		
3	6	5.411	4.732	3.129	5.882	5.126	3		
3	7	4.322	4.851	3.355	5.843	4.724	3		
3	8	4.627	3.906	2.939	3.970	3.945	3		
3	9	3.464	3.345	3.005	2.453	3.704	4		
3	10	4.734	4.784	2.616	3.370	4.035	3		
								9	
4	1	2.945	1.768	3.714	4.417	4.632	2		
4	2	5.270	5.109	4.394	2.270	3.286	4		
4	3	7.188	6.497	6.860	5.724	6.000	4		
4	4	3.524	3.152	3.407	3.028	3.715	4		
4	5	3.611	3.917	4.081	3.536	4.876	4		
4	6	3.608	3.675	3.819	2.475	3.976	4		
4	7	3.569	3.024	4.106	3.003	4.200	4		
4	8	6.038	5.217	5.776	4.481	5.015	4		
4	9	5.485	5.209	4.782	2.206	3.842	4		
4	10	6.633	6.464	5.638	4.230	5.854	4		
4	11	3.739	3.574	4.019	2.612	3.273	4		
4	12	4.037	4.595	4.600	3.708	4.127	4		
4	13	4.277	3.767	4.318	3.645	4.831	4		
4	14	8.326	8.621	8.330	6.623	5.998	5		
								12	
5	1	3.614	3.356	2.809	5.292	4.558	3		
5	2	7.644	8.340	7.692	7.290	5.862	5		
5	3	4.252	3.893	2.827	3.548	3.040	3		
5	4	5.642	5.085	3.330	4.970	4.154	3		
5	5	4.441	3.783	2.466	3.849	3.331	3		
5	6	4.502	3.574	2.823	3.856	3.747	3		
5	7	4.227	4.201	3.282	4.691	3.148	5		
5	8	5.685	4.505	3.688	5.002	5.583	3		
5	9	5.605	5.081	3.865	4.528	5.098	3		
5	10	4.581	3.196	2.576	4.218	4.627	3		
								2	33

DECISION FUNCTION VALUES (DIST. TO LANG. MEAN) (TRAINING DATA)

FEATURES USED: 6 7 8 9 10 11 13 17 20 21 22 24 27 28 33 34

LNG	SPKR	L1	L2	L3	L4	L5	DEC.	# CORR.	TOT.
		LANGUAGE							
1	1	2.298	3.450	4.005	5.600	5.308	1		
1	2	2.451	4.710	4.135	5.133	4.295	1		
1	3	2.538	3.397	3.828	3.621	4.101	1		
1	4	3.094	3.746	3.561	3.365	3.556	1		
1	5	1.562	3.431	3.117	4.595	4.212	1		
1	6	3.367	5.930	5.034	7.404	6.866	1		
1	7	2.292	3.428	3.604	4.838	4.796	1		
1	8	2.844	4.287	4.545	6.354	6.065	1		
1	9	3.699	5.638	5.342	6.560	5.575	1		
1	10	3.192	4.050	4.926	5.093	5.531	1		
								10	
2	1	4.220	2.366	4.148	4.998	4.730	2		
2	2	4.703	2.039	4.456	4.220	4.517	2		
2	3	4.553	3.040	4.502	4.416	3.291	2		
2	4	4.100	3.497	5.552	6.410	6.153	2		
2	5	3.132	2.758	2.867	5.260	4.597	2		
2	6	4.601	2.367	4.674	6.140	5.578	2		
2	7	4.320	2.005	4.226	4.486	4.410	2		
2	8	3.939	3.665	4.120	6.094	5.292	2		
2	9	5.587	4.115	4.602	3.907	4.666	4		
2	10	3.736	1.536	3.239	4.126	4.415	2		
								9	
3	1	5.534	5.567	3.416	6.122	6.135	3		
3	2	3.111	3.943	2.367	4.316	3.024	3		
3	3	4.959	4.956	2.786	4.249	5.089	3		
3	4	4.071	4.159	2.946	5.863	5.285	3		
3	5	3.580	4.451	2.783	5.877	5.125	3		
3	6	3.256	2.683	2.628	3.161	3.442	3		
3	7	3.990	3.242	1.967	4.974	4.645	3		
3	8	4.414	3.946	2.241	4.882	4.157	3		
3	9	6.760	7.096	5.369	5.994	5.433	3		
3	10	3.678	3.007	3.358	4.187	4.302	2		
								9	
4	1	5.895	6.001	5.712	2.889	4.076	4		
4	2	4.947	4.507	3.930	3.786	5.129	4		
4	3	4.657	4.276	3.634	2.555	4.332	4		
4	4	5.947	5.476	5.306	2.508	3.737	4		
4	5	3.816	3.618	3.869	2.182	3.195	4		
4	6	5.753	5.388	4.997	1.805	3.160	4		
4	7	7.022	6.893	6.678	3.535	5.685	4		
4	8	5.129	3.985	4.233	3.095	3.332	4		
4	9	5.354	4.896	4.610	1.995	2.449	4		
4	10	4.460	4.312	4.494	2.450	3.052	4		
								10	
5	1	3.523	3.805	3.318	3.460	2.732	5		
5	2	3.439	3.000	3.243	2.192	2.081	5		
5	3	4.316	4.831	4.221	3.257	2.806	5		
5	4	10.902	10.863	10.336	8.861	7.552	5		
5	5	6.414	5.455	6.170	4.174	2.941	5		
5	6	6.083	5.840	5.470	3.814	3.308	5		
5	7	3.801	3.554	4.399	3.699	3.187	5		
5	8	4.929	5.344	4.729	3.719	2.289	5		
5	9	5.480	5.016	4.551	5.965	4.031	5		
5	10	6.849	5.187	4.726	6.979	6.539	3		
								9	47

DECISION FUNCTION VALUES (DIST. TO LANG. MEAN) (TEST DATA)

FEATURES USED: 6 7 8 9 10 11 13 17 20 21 22 24 27 28 33 34

		LANGUAGE							
LNG	SPKR	L1	L2	L3	L4	L5	DEC.	# CORR.	TOT.
1	1	3.493	5.942	5.714	7.270	6.762	1		
1	2	3.652	3.750	3.473	3.879	4.430	3		
1	3	2.772	3.537	3.337	3.768	3.461	1		
1	4	3.642	3.774	3.499	3.302	3.718	4		
1	5	2.614	4.551	4.809	6.385	5.460	1		
1	6	3.434	4.907	4.728	4.563	4.559	1		
1	7	5.263	5.423	5.331	3.295	3.955	4		
1	8	3.872	4.298	3.853	2.996	3.516	4		
1	9	3.510	3.924	3.132	3.317	3.402	3		
1	10	3.022	3.480	3.780	3.596	4.092	1		
								5	
2	1	2.941	3.529	2.415	4.613	3.587	3		
2	2	7.486	6.408	6.421	4.892	6.041	4		
2	3	3.382	4.242	4.644	4.544	4.643	1		
2	4	8.381	8.056	7.387	5.624	4.986	5		
2	5	4.185	3.870	3.363	3.351	2.783	5		
2	6	4.054	2.864	3.447	3.020	3.602	2		
								1	
3	1	4.253	3.128	1.834	4.478	4.096	3		
3	2	4.284	3.042	2.242	4.413	4.329	3		
3	3	5.017	3.923	2.582	5.001	4.926	3		
3	4	3.589	3.322	2.038	4.569	4.621	3		
3	5	5.614	6.007	4.519	7.023	6.211	3		
3	6	5.640	4.830	3.235	6.124	5.218	3		
3	7	4.414	4.933	3.379	6.116	4.811	3		
3	8	5.108	4.194	3.197	4.025	3.713	3		
3	9	4.040	3.534	3.238	2.573	3.580	4		
3	10	5.103	4.930	2.760	3.619	4.100	3		
								9	
4	1	2.938	2.017	3.841	5.014	4.914	2		
4	2	6.024	5.560	4.883	2.417	3.270	4		
4	3	8.291	7.322	7.764	6.140	6.494	4		
4	4	4.323	3.495	3.896	3.229	3.830	4		
4	5	3.937	3.992	4.141	3.760	4.812	4		
4	6	3.742	3.630	3.810	2.996	4.231	4		
4	7	4.284	3.326	4.464	2.984	3.987	4		
4	8	6.666	5.875	6.343	4.665	5.018	4		
4	9	7.523	6.846	6.779	4.100	5.291	4		
4	10	7.213	6.837	5.999	4.285	5.702	4		
4	11	4.436	3.818	4.321	2.684	3.313	4		
4	12	3.916	4.365	4.337	3.830	4.273	4		
4	13	4.738	3.964	4.470	3.538	4.492	4		
4	14	8.292	8.244	8.010	6.004	5.751	5		
								12	
5	1	4.292	3.569	3.025	5.532	4.676	3		
5	2	6.312	6.529	5.731	5.644	4.579	5		
5	3	4.687	4.069	2.996	3.641	2.996	5		
5	4	6.000	5.226	3.468	5.283	4.510	3		
5	5	4.656	3.977	2.668	3.944	3.244	3		
5	6	5.081	3.998	3.178	3.979	3.556	3		
5	7	4.623	4.348	3.399	4.836	3.138	5		
5	8	5.968	4.620	3.813	5.192	5.588	3		
5	9	5.646	5.125	3.932	4.653	5.113	3		
5	10	4.955	3.354	2.744	4.118	4.249	3		
								3	30

REFERENCES

1. R. G. Leonard and G. R. Doddington, "Automatic Language Identification," Final Report, RADC-TR-74-200, August 1974 (AD785397).

2. R. G. Leonard and G. R. Doddington, "Automatic Classification of Languages," Final Report, RADC-TR-75-264, October 1975. (AD B008 708L)

3. G. R. Doddington, R. E. Helms, and B. M. Hydrick, "Speaker Verification III," Final Technical Report, RADC-TR-76-262, August 1976. (AD B014 720L)

MISSION
of
Rome Air Development Center

RADC plans and conducts research, exploratory and advanced development programs in command, control, and communications (C^3) activities, and in the C^3 areas of information sciences and intelligence. 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.

AMERICAN REVOLUTION BICENTENNIAL 1776-1976