1984

Algorithms for simultaneous speaker verification and digit recognition.

Aladangady. Udaya
University of Windsor

THE FEASIBILITY OF APPLYING VIBRATION MONITORING TECHNIQUES TO HIGH VOLUME MULTISTATION TRANSFER MACHINES

by

SAUW-YOENG TJONG

A Thesis
Submitted to the
Faculty of Graduate Studies and Research
through the Department of Mechanical Engineering in
Partial Fulfillment of the Requirements for the
Degree of Master of Applied Science
at the University of Windsor

Windsor, Ontario, Canada
1984

ABSTRACT

Experiments investigating simultaneous Automatic Message and Speaker Recognition (AMSR) are reported in this thesis. Several different parameter sets and their subsets, derived from input speech were examined for their effectiveness for AMSR realization. The development of a high accuracy AMSR system based on new feature sets with new similarity measures is discussed. The first stage of experiments deals with the evaluation of combined speaker verification and digit recognition based on single digit utterances. A known technique of automatic speech recognition is examined for its effectiveness for combined speaker and digit recognition. New techniques based on orthogonal parameters derived from different features, with two similarity measures, also are investigated for AMSR realization. The second stage of experiments utilizes spoken digit strings to obtain significantly higher AMSR accuracies than is possible with any single digit utterance. For completeness of discussion a review of relevant literature, data collection, front-end processing and feature extraction functions are also presented.

ACKNOWLEDGEMENT

The author expresses his sincere gratitude towards his supervisor Dr. M. Shridhar for all the valuable discussions, suggestions and constructive criticism throughout the study period. Without his active guidance and support, this work could not have accomplished. The valuable advice from Dr. N. Mohankrishnan is also gratefully acknowledged.

Thanks are due to Professor P. H. Alexander, Dr. Dan Watt, Dr. M. Shridhar, Mr. John Robinson, Mr. Ed Klinard, Mr. Joe Novosad, Mr. Alan Thibert and Dr. Mohankrishnan for contributing their speech for this research.

I would like to express my sincerest thanks and gratitude to my wife, Suman, for her constant encouragement. To my mother and father, I extend my sincere thanks for their constant moral support during this study and research period.

Thanks is due to all other faculty members and all the graduate students of this department for their constant assistances.

LIST OF CONTENTS

Appendix — page

ABSTRACT

A study was undertaken to determine the feasibility of applying vibration monitoring techniques to high volume multistation transfer machines.

Recent published literature on machinery health monitoring is reviewed with special emphasis on vibration monitoring. A complete bibliography of 255 references is appended, together with summary chart, in which the subject is classified by topics.

A field study was undertaken to determine the feasibility of applying vibration monitoring techniques to high volume, multistation transfer machines installed in one of the leading automotive engine plants. An accelerometer and a tape recorder were used to obtain the vibration data. It was shown that repeatable vibration measurements were possible under "in plant" conditions and that future trends in both the overall and spectral acceleration levels were readily apparent. Furthermore, for one particular machining station, prediction of bearing failure was documented.

As a result of the successful "in plant" manual vibration monitoring, a series of controlled bearing failure tests were performed in order to determine the most suitable vibration analysis technique for identifying specific types of failure. The results of defects which were induced to

LIST OF ILLUSTRATIONS

LIST OF APPENDICES

Page

recognition systems were dedicated to industrial applications in which user's hands and eyes were already busy with their normal work requirements. Other successful applications for voice input systems are the following :

Automatic material handling, automatic quality control and inspection, voiced programming of numerically controlled machines, voice actuated wheel chairs, voice data entry into computers, cartography and defense mapping, airplane cockpit communications. Some of these applications are shown in the figure 1.1.

Figure		Page
5.2b	The Shaft and Bearing Assembly	111
5.3	The Spindle Layout Including All the Measuring Positions	112
5.4	The Schematic of Experimental Set Up	113
5.5	An Induced Outer Race Defect	117
5.6	An Induced Ball Defect	117
5.7	An Induced Inner Race Defect	118
5.8	Induced Multiple Defects	118
5.9a/b	Comparison of Frequency Spectra of Prior to Removal and After Being Removed and Reassembled	124
5.10a/b	Comparison of Frequency Spectra of Incorrectly Installed Bearing and Correctly Installed Bearing	125
5.11a	Acceleration Versus Time for Outer Race Defect (Position 5B)	127
5.11b	Outer Race Defect Spectrum Versus Baseline Spectrum for Position 5B, 0 - 1.6 kHz	127
5.11c	Outer Race Defect Spectrum Versus Baseline Spectrum for Position 5B (Decibel Scale), 0 - 1.6 kHz	128
5.11d	Outer Race Defect Spectrum Versus Baseline Spectrum for Position 5B, 0 - 25.6 kHz	128
5.12a	Acceleration Versus Time for Ball Defect (Position 6B)	132
5.12b	Ball Defect Spectrum Versus Baseline Spectrum for Position 6B, 0 - 1.6 kHz	132
5.12c	Ball Defect Spectrum Versus Baseline Spectrum for Position 6B (Decibel Scale), 0 - 1.6 kHz	133

LIST OF TABLES

NOMENCLATURE

A/D	analog to digital
ADC	analog to digital converter
AM	amplitude modulation
A_n	the average after n time records
A_{n-1}	the average after (n-1) time records
A Spec	auto spectrum
BW	bandwidth
cm	centimeter
cpm	cycle per minute
d_B	the ball diameter
dB	decibel
D_I	the inner race contact diameter
D_O	the outer race contact diameter
D_P	the pitch diameter
DR	direct
f_A	the ball assembly (fundamental train) frequency
f_B	the ball spin frequency
f_{ht}	the highest frequency of the transition band
f_I	the ball pass frequency of the inner race
f_{in}	the input frequency
f_{max}	the maximum frequency of interest
f_{min}	the minimum frequency of interest

f_O	the ball pass frequency of the outer race
f_R	the rotational shaft frequency
f_S	the frequency of the ADC sampling operation
FFT	fast fourier transform
FM	frequency modulation
g	unit of acceleration (9.81 meter / second ^{2})
g-SE	unit of acceleration of spike energy
Hz	Hertz
in	inch
I_i	the i^{th} time record
I_n	the n^{th} time record
IRD	Indusrial Research and Development Corporation
ISO	International Standards Organization
kHz	kiloHertz
K	the number of samples in a time record
ℓ_B	the linear travel of ball center
ℓ_I	the linear travel of inner race
ℓ_O	the linear travel of ball on the outer race
L.	left
LPF	low pass filter
mil	0.001 inch
ms	millisecond
MAG	magnitude
n	the number of time records
N	the decay constant

N_B	the number of balls
rms	root mean square
rpm	revolution per minute
R.	right
R #	reading number
sec	second
SE	spike energy
STA.	station
T	the total time record length
V_B	the linear velocity of the ball
V_I	the linear velocity of the inner race
#A	the number of averages
β	the angle change of the inner race
$\emptyset$	the contact angle
α_I	the angular travel of inner race
α_o	the angular travel of ball center
Δt	the time interval between digital history values
Δf	the frequency resolution
Σ	the summation

I. INTRODUCTION

Today, achieving higher productivity is the single most important goal of machine tool builders. This, however, requires not only innovative concepts in the part transfer and metal removal processes, but also the integration of monitoring and diagnostic tools that permit early warning of machine component failure. Vibration monitoring is one of the main techniques used to predict and diagnose a wide range of incipient failures in rotating machines. Such capabilities would substantially reduce the problem of unscheduled maintenance, minimize additional damage to the machine, permit advanced planning of changes in production schedules, reduce spare parts inventory and return the machine to operating condition quickly.

Therefore, it is obvious that machine tool builders who wish to succeed in the highly competitive market place of the future, must plan today for the integration of vibration monitoring systems into their machines.

A transfer machine line is a collection of automatic machining stations of all types (see Figure 1.1). The workpiece, such as an engine block, enters the transfer line as a rough casting and leaves it completely machined at a rate of approximately 200 units per hour. The rate of production is very high, thus unscheduled machine downtime

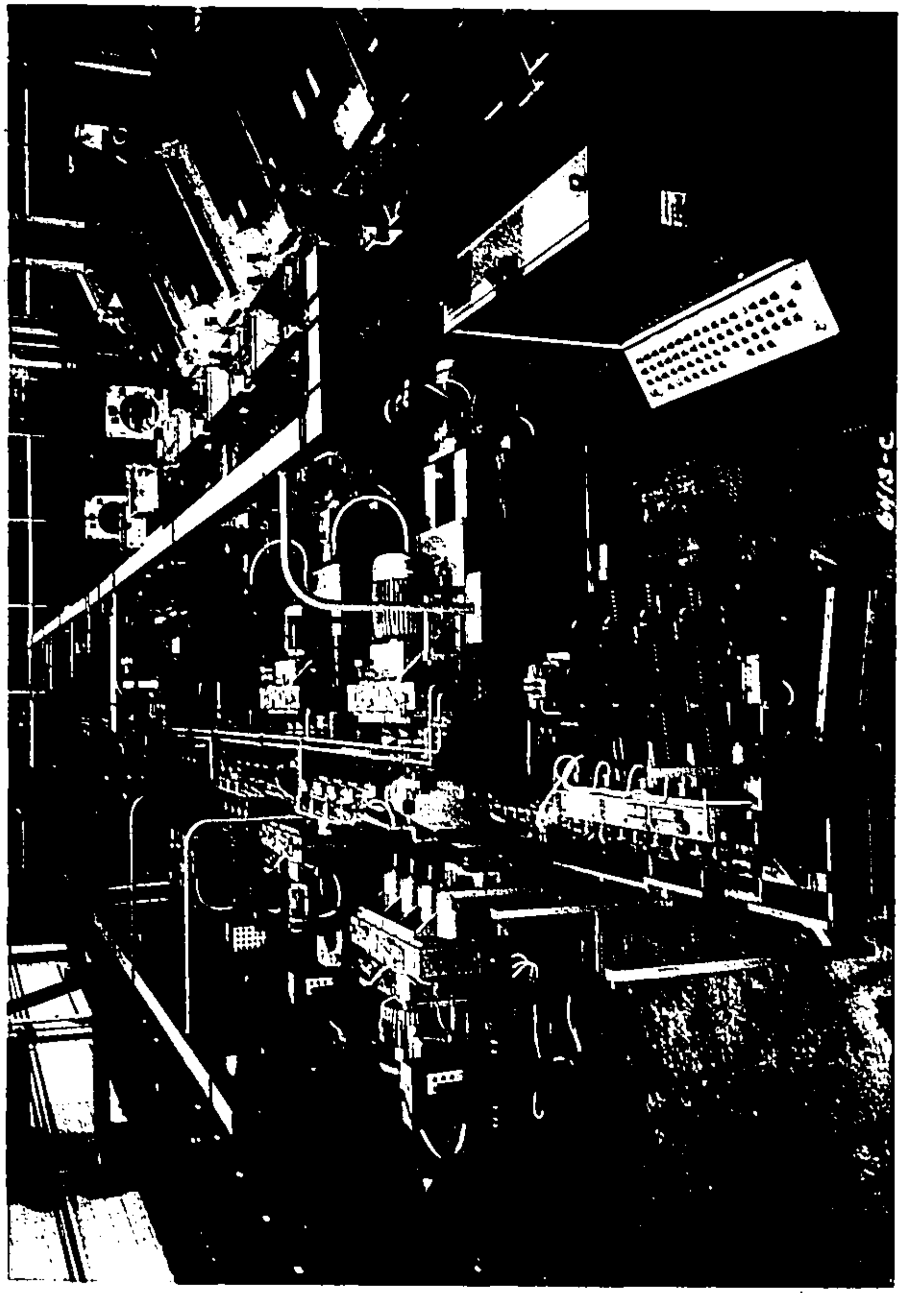


FIGURE 1.1: TYPICAL HIGH VOLUME MULTISTATION TRANSFER MACHINE

can significantly reduce productivity. Furthermore, the failure of one machining station in a transfer line will automatically stall the whole line production. Bearing failures are at present the most common causes of machine downtime. Every bearing has a limited life which is strongly influenced by the method of installation, operating conditions and maintenance received. Thus the reliability, efficiency and safety of the spindles used in the machining station depend on bearings functioning properly.

In view of these considerations the following objectives were set for this thesis:

a. To review the recently published literature on machinery condition monitoring with special emphasis on vibration monitoring techniques.

b. To determine the feasibility of applying vibration monitoring techniques on the high volume, multi-station transfer machines installed in the engine plant of a leading automotive manufacturer.

c. To obtain representative overall acceleration levels as well as frequency spectra from designated machining operations for use as a "baseline" in a future vibration monitoring system.

d. To identify induced bearing defects in a typical single spindle machining station using time domain analysis as well as the frequency domain analysis.

e. To summarize the results and recommend areas of future research and development.

II. LITERATURE SURVEY

The subject of machine health monitoring has long been a widely-documented field in machinery research. In fact, current published literature focuses largely on this subject as an important tool in extensive machine studies. This particular review concentrates on vibration monitoring with special emphasis on bearing and gear failures and automated (computerized) vibration monitoring systems. The topics reviewed include the sources of vibration in rotating machinery, instrumentation, measurement techniques, data processing techniques, applications to various types of machinery and systems, and users' experiences.

Because textbooks offer a very limited amount of practical information on vibration monitoring and analysis of rotating machinery, this review has predominantly made use of periodicals, proceedings and seminar notes. A complete bibliography of 255 references is given in alphabetical order in Appendix A. A summary chart of the full bibliography with classification of technical papers by topics is given in Appendix B. Of the references listed in Appendix A, 155 have been reviewed in detail and are referred directly in this thesis. These entries are embodied in a separate "References" section.

2.1.3.1 DESCRIPTION OF THE ALGORITHM

The samples corresponding to background noise from the first 100 msecs of a recording were utilized for calculating statistical measurements of background silence. These measurements include the average zero-crossing rate (IZC'), standard deviation of zero-crossing rate (σ) and peak energy (IMN). A zero-crossing threshold, IZCT, to discriminate silence from speech region was calculated as follows:

$$IZCT = MIN\ (IF\ ,\ IZC' + 2\sigma)$$

where IF is a fixed threshold of 25 zero-crossings per 10 msecs sampling interval.

The energy feature for the speech region was calculated and its peak energy IMX along with energy statistics of the background noise were used to calculate two thresholds ITL and ITU according to the rule :

$$I1 = 0.03 * (IMX - IMN) + IMN$$

where I1 is a level at 3% of the peak energy.

$$I2 = 4 * IMN$$

where I2 is a level set at four times the silence energy.

$$ITL = MIN\ (I1\ ,\ I2)$$

where ITL, lower energy threshold, is the minimum of two thresholds I1 & I2.

$$ITU = 5 * ITL$$

Where ITU is ~~the~~ upper energy threshold which is 5 times the lower threshold. Upper and lower energy thresholds ITL and ITU were used to roughly estimate the begin and end points of speech segment. Later a back-tracking from the rough estimate of begin point was made to verify if samples exceed the ZCR threshold IZCT. This is to detect any low energy, high frequency fricative sounds in the beginning of the utterance. Similarly a forward tracking from the initial rough estimate of the end point was made to verify the presence of weak fricatives at the end of the utterance. A detailed structure and details of the algorithm is furnished in the Appendix [A].

This algorithm for word end point detection has given reasonably good results over ten digit vocabulary spoken by seven speakers.

2.1.4 WINDOWING OF SPEECH SAMPLES

Windows are weighting functions applied to pre-emphasized, word endpoint detected data. According to F.J.Harris [12], windowing helps in accomplishing the following :

1. Reduction of spectral leakage associated with finite observation intervals.
2. Reduction of the order of discontinuity at the boundary of the periodic extension.

Windowing was accomplished by multiplicative weighting to the input data in discrete time-domain and is as given below :

$$h(n) = 1 \qquad 0 \leqslant n \leqslant N-1$$
$$= 0 \qquad \text{otherwise}$$

where h(n) is the weighting function and N is the number of samples in the windowed segment.

This type of weighting is known as rectangular window because it gives equal weightage of 1 inside the window interval. From frequency domain point of view this type of sharp transition of weightings at the begin and end of the window will give rise to ringing in the spectrum thereby causing aliasing errors. A better weighting function to smoothly bring the weighting function to zero at boundaries of the window interval is given by :

$$h(n) = 0.54-0.46*COS(2*PI*n/(N-1)) \qquad \text{for } 0 \leqslant n \leqslant N-1$$
$$= 0 \qquad \text{otherwise}$$

This window function is known as Hamming window and was used for windowing speech while evaluating frequency-domain feature sets. Rectangular Windowing was used while calculating time-domain feature sets.

The windowed speech samples in time-domain are mathematically represented as follows :

$$s''(n) = s(n) * h(n)$$
$$= \sum_{m=-\infty}^{\infty} s(m)\ h(n-m)$$

where * indicates discrete convolution. $s(n)$ is n th input speech sample and $s''(n)$ is n th windowed speech sample.

After sampling, digitizing, pre-emphasis and word end point detection, pre-processed speech samples within the end points were stored in the digital mag-tape storage. This data stored in the mag-tapes, form the input to the next crucial processing phase known as feature extraction.

Chapter III

FEATURE EXTRACTION

Significant data compression and computational savings can be accomplished if a few representative features extracted from speech carry all essential and relevant information contained in input speech. A set of representatives in the form of different parameter sets obtained as a result of time-domain or frequency-domain analysis were tested for their speech sensitive and talker sensitive characteristics. A list of different feature sets used in these experimental investigations for AMSR system are explained here.

The feature sets known as energy, zero-crossing rate, normalized error and pole frequencies of 2-pole linear prediction analysis and first two formants were demonstrated by Sambur and Rabiner [3] to be message sensitive parameters of speech. Also Linear Predictor Coefficients (LPCs) were demonstrated by Itakura [4], to be message sensitive for a designated speaker. Later Atal [13] and Sambur [5] demonstrated the effectiveness of orthogonal LPCs for automatic speaker recognition. All these parameters including two new sets of parameters known as Direct Fourier Transform of speech (DFTs) and Inverse Filter Spectral Coefficients

(IFSCs), were investigated in these experimental investigations for an efficient AMSR system.

The pre-processed data corresponding to each vocabulary entry was read from mag-tapes and placed in disk storage and parameter extraction was carried out in windowed segments. The concept of short-time feature extraction (feature extraction from windowed segments) is fundamental for describing quasi-stationary (slowly time varying) signal such as speech.

3.1 TIME-DOMAIN FEATURES

3.1.1 ENERGY FUNCTION

This parameter provides a representation that reflects speech waveform amplitude variations and is defined as sum of the squared values of speech samples in a given windowed segment.

$$E(n) = \sum_{m=-\infty}^{\infty} x^2(m)\; h(n-m)$$

where h(n) is rectangular window.

Another definition of energy function is the sum of absolute values of windowed speech samples and is given by :

$$E(n) = \sum_{m=-\infty}^{\infty} \left| s(n)\; h(n-m) \right|$$

where h(n) is rectangular window.

3.1.2 LINEAR PREDICTION COEFFICIENTS (LPCs)

A very popular and effective characterization of speech and speaker was realized through the use of a linear discrete model defined by the transfer function :

$$H(z) = G \,/\, (1 - \sum_{i=1}^{P} a_i z^{-i})$$
$$= G \,/\, A(z)$$

Where G is the gain of the model

a_i's are LPCs

A(z) is the inverse filter transfer function

P is the order of the LPC model.

An equivalent time-domain description is obtained as :

$$s(n) = \sum_{i=1}^{p} a_i s(n-i) + G\ e(n)$$

where s(n) is the predicted sample at n th instant.

s(n-1) is the predicted sample at (n-1) th instant.

e(n) is the glottal excitation at n th instant.

a_i's are Linear Prediction Coefficients (LPCs).

Considering that neither vocal tract shape nor the glottal waveform changes significantly over 24 msecs, the LPC measurements were conducted on 24 msec Hamming windowed segments. The predictor coefficients were determined by minimizing the mean squared prediction error between actual and predicted speech samples as given below :

$$
\begin{aligned}
E &= \sum_{n} (s(n) - s'(n)) \\
&= \sum_{n} (s(n) - \sum_{k=1}^{p} a_k s(n-k)) \\
&= \sum_{n} (\sum_{k=0}^{p} a_k s(n-k))
\end{aligned}
$$

A set of linear equations is obtained by taking partial derivatives of this squared error with respect to the a 's and equating to zero. Solution of this set of equations give the LPCs. Details of algorithm for solving this set of equations are given in Appendix [B].

3.1.3 ZERO CROSSING RATE (ZCR)

This measurement is perhaps the simplest method of estimating the signal's amplitude spectrum. This parameter calculation consists of counting the number of times the voltage analogue of the signal changes algebraic sign (from plus to minus or from minus to plus) in an analysis segment. Mathematically this is represented as follows :

$$z(n) = \sum_{m=-\infty}^{\infty} \left| \mathrm{sgn}[x(m)] - \mathrm{sgn}[x(m-1)] \right| w(n-m)$$

where $\mathrm{sgn}[x(n)] = 1 \quad \text{for} \quad x(n) \geqslant 0$

$\qquad = -1 \quad \text{for} \quad x(n) < 0$

and $\quad w(n) = 1/2N \quad \text{for} \quad 0 \leqslant n \leqslant N-1$

$\qquad = 0 \quad \text{otherwise}$

where N is the total number of samples in the analysis segment.

Zero-crossing rate measurement is economically attractive because it can be accomplished by ~~using~~ simple electronic devices.

3.1.4 NORMALIZED ERROR

This feature set is obtained from a two-pole LPC analysis on a given segment of speech. The squared prediction error between actual and predicted speech samples is known as normalized error. This is mathematically represented as :

$$\sum_{n=n_0}^{n_1} [e(n)^2] = \sum_{n=n0}^{n1} (s(n)-s'(n))^2 = \sum_{n=n0}^{n1} \Big(s(n)-\sum_{k=1}^{P} a_k s(n-k)\Big)^2$$

$$= \sum_{n=n0}^{n1} \Big(\sum_{k=0}^{p} s(n-k)\ a_k \Big)^2$$

where a_k's are calculated as explained in the Appendix [B].

This parameter set contains important information about the spread of spectral energy in a given utterance.

3.1.5 POLE FREQUENCY

This feature set is obtained from a two-pole LPC analysis of a given utterance. A characterization of speech realized through the use of a 2-pole LPC model is given by :

$$H(z) = 1 / (1 - \sum_{i=1}^{p} a_i z^{-i})$$

and its frequency response is given by :

$$H(e^{jwT}) = 1 / (1 - a1\ e^{-jwT} - a2\ e^{-j2wT})$$

Where a1, a2 are LPCs calculated as given in Appendix [B]

T is the sampling period (0.1 msecs).

Two poles of this transfer function are complex quantities given by :

$$z1 , z2 = (r1+jw1), (r2+jw2)$$

where w1 and w2 are known as pole frequencies(in radians/sec).

This feature set essentially distinguishes high frequency type sounds from low frequency type sounds in an utterance.

3.2 FREQUENCY DOMAIN FEATURES

3.2.1 FORMANTS

The speech waveform can be modeled as the response of a resonator (the vocal tract) to a series of pulses (quasi periodic glottal pulses during voiced sounds, or noise generated at constriction during unvoiced sounds). The resonances of the vocal tract are called as formants, and they are manifested in the spectral domain by energy maxima at the resonant frequencies. The formant frequencies are an important cue in the characterization of speech sounds, and therefore an automatic algorithm for reliably computing these frequencies would be useful for speech recognition research. One approach to the problem is by peak-picking from spectral pattern of impulse response of the inverse filter of the LPC model. It has been found from experimental observations by Stephen S. McCandless [14], that a minimum of 14 LPCs are required to eliminate the chances of merging of adjacent formant peaks of certain sounds.

Once the LPC coefficients a_k's are available, it is easy to obtain the approximated spectrum of s(n). Evaluation of the magnitude of the transfer function H(z) of the filter represented by the coefficients a_k's at N equally spaced samples along the unit circle in the z-plane.

$$H(z) = 1 / (1 - \sum_{k=1}^{14} a_k z^{-k})$$

where H(z) is evaluated at z = exp[j(2*PI*n/N)] for n=0,1,..,N-1

N can be chosen arbitrarily large to increase frequency resolution, at the expense of computation time. A sequence of 128 complex points ($1, a_k$: k=1,..,14, appended zeroes) is formed. Fourier transform calculations are performed on this discrete input sequence by means of fast Fourier transform algorithm. Magnitudes (in dBs) of the first 64 DFT points is calculated. A plot of these magnitude points versus frequency, show a number of different peaks in the frequency response. Center frequencies of different peaks are known as formant frequencies.

3.2.2 INVERSE FILTER SPECTRAL COEFFICIENTS (IFSCs)

The speech production model of LPC analysis provides a system function :

$$H(z) = 1 / (1 - \sum_{k=1}^{P} a_k z^{-k}) = 1 / A(z) ;$$

A (z) is known as an inverse filter and

$$s(n) - \sum_{k=1}^{P} a_k s(n-k) = e(n),$$

where e (n) is the glottal excitation; the a_k's are LPCs; P is the order of the LPC model.

The frequency response of the Inverse filter is given by :

$$A(e^{j\omega T}) = (1 - \sum_{k=1}^{P} a_k e^{-j\omega Tk})$$

$A(e^{j\omega T})$ is the discrete Fourier transform of sequence $(1, a_0, a_1, \ldots, a_P)$. For instance for a 12 th order LPC model a 64 length sequence can be formed (1, 12 LPC coefficients and 51 appended zeroes) and Fourier analysis is performed by employing Fast Fourier Transform algorithm. Magnitudes of the complex DFT sequence is calculated. The magnitude points on one side of the maximum are mirror images of those on the other side of the maximum. The first 32 magnitude points form inverse filter spectral coefficients (IFSCs) for the analysis segment.

3.2.3 DIRECT FOURIER TRANSFORM OF SPEECH (DFTs)

Direct Fourier transform calculation on speech samples is an efficient method for estimating its amplitude spectrum. Just like the filter bank method (where average energy in different frequency bands was required), the Fast

Fourier Transform provides a computationally efficient method of estimating the average amplitude spectrum at different discrete frequency points.

Preprocessed speech samples are input to the Fast Fourier Transform algorithm. 240 speech samples with 16 appended zeroes form a 256 sample length discrete input sequence for the FFT algorithm. The magnitude of the output sequence of complex direct Fourier transform coefficients are calculated. DFT representations are periodic with modulo-256, hence only first 128 magnitude points of the DFT sequence are considered for further calculations. Arithmetic averages of four consecutive magnitude points of the DFT sequence give 32 average values. These 32 discrete numbers per analysis segment are known as the Direct Fourier Transform of speech or DFTs.

3.3 ORTHOGONAL PARAMETERS

By means of eigenvalues and eigenvectors of a matrix of covariances of measurements made on a population of utterances from a given talker, a set of orthogonal parameters are generated. These orthogonal parameters can be derived from any set of features viz: LPCs, inverse filter spectral coefficients, direct Fourier transform of speech, parcor coefficients, log area coefficients. Orthogonalization of parameter sets essentially make these parameters mutually uncorrelated thereby bringing out latent speech sensitive, speaker sensitive or medium sensitive characteristics from them. Different steps involved in calculation of orthogonal parameters are as given below :

1. Let x_{ij} : $i=1,2,\ldots,M$; $j=1,2,\ldots,NF$ be the parameter set where x_{ij} is the i th parameter of the j th frame, M is the number of parameters in the set and NF is the total number of analysis frames in the utterance.

2. Compute the covariance matrix [C] of the parameter set, where [C] : c_{lm} : $l=1,2,\ldots,M$; $m=1,2,\ldots,M$ is given by :

$$C_{lm} = 1/(NF-1)\sum_{j=1}^{NF}(x_{lj}-\bar{x}_l)(x_{mj}-\bar{x}_m)$$

and

$$\bar{x}_l = (1/NF)\sum_{j=1}^{NF} x_{lj}$$ is the average value of the l th parameter.

3. Compute the eigen values λ_l : $l=1,2,\ldots,M$ and the eigen vectors T_l of the matrix [C] by solving $|C-\lambda I|=0$ for λ_l"s and by solving $CT_l=\lambda_l T_l$ for T_l .

4. Normalize T_l to unit length.

5. Evaluate the orthogonal parameters (ϕ_{ij} : $i=1,2,\ldots,M$; $j=1,2,\ldots,NF$) as follows :

$$\phi_{ij} = \sum_{l=1}^{M} t_{il}\, x_{lj}$$

where ϕ_{ij} is the i th orthogonal parameter in the j th frame; t_{il} is the l th element of the i th eigen vector T_i .

These steps of processing form speaker reference orthogonal parameters, as they are based on covariance of several utterances of that speaker. To generate test orthogonal parameters, a dot product of reference eigen vectors with test parameter set is to be evaluated. For generating overall average orthogonal parameters (for test data), a dot product of reference eigen vectors and overall average orthogonal parameter set is to be evaluated. To obtain framewise orthogonal parameters (for test data), a dot product of reference eigen vectors and framewise parameter sets are to be calculated.

Chapter IV

EXPERIMENTAL INVESTIGATIONS FOR AMSR SYSTEM

4.1 PREVIEW OF THE EXPERIMENTS

The target of this investigation was to locate a feasible method for recognizing a speaker and his spoken message (spoken digits) simultaneously. This problem can be attacked in two different ways :

1. To recognize the spoken message (or word) in speaker-independent mode followed by speaker recognition in text-dependent mode.
2. To recognize the speaker in text-independent mode, followed by message (or word) recognition in speaker-dependent mode.

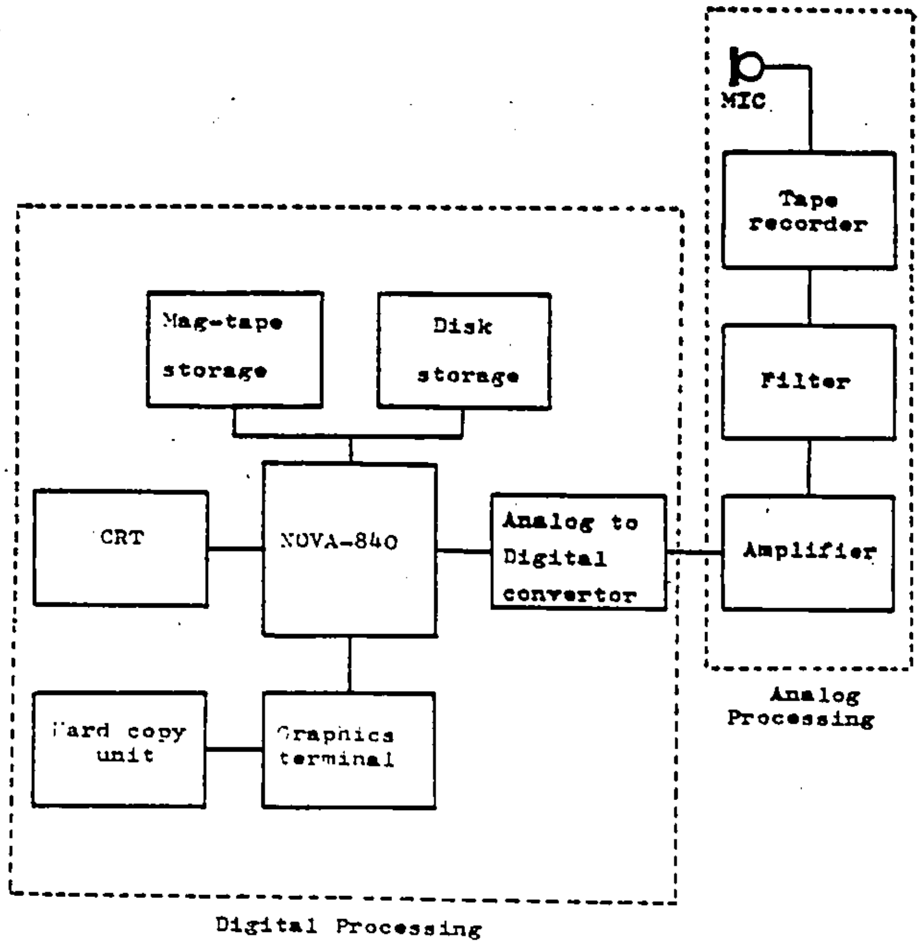
Exhaustive tests were conducted with a known digit recognition technique based on energy, zero-crossing rate, normalized error (of 2-pole LPC model), pole frequencies (from 2-pole LPC model), and formants. These features were inherently speech sensitive, thereby giving a speaker-independent nature to this scheme of recognition. As this approach is ineffective for automatic speaker recognition, simultaneous recognition of speech and speaker cannot be attempted with these sets of features. The experimental investigations on this speaker-independent speech recognition did not contribute to the target of this thesis.

First section deals with descriptions of experiments conducted with a speaker-dependent word recognition system based on 12 th order LPC model. This approach utilizes dynamic programming, time warping and Itakura's similarity measure for speaker and digit recognition. Second section contains the descriptions of experiments conducted with orthogonal parameters (LPCs, DFTs, IFSCs) for speaker verification (both text-dependent and text-independent modes) and isolated digit recognition. Also the above method was investigated with two similarity measures 'average distance' and 'distance of averages'. The experiments were extended with different subsets of orthogonal LPC parameters derived from single digit utterances and the relevant results are furnished in this section. Third section consists of descriptions of experiments conducted with sequences of spoken digits (3 digit strings, 7 digit strings) with orthogonal parameter (LPCs, DFTs) analysis. Average distance and distance of averages computed across digit string utterances in orthogonal parameter domain, also are investigated. Descriptions of these experiments and test results of orthogonal parameter subsets for simultaneous speaker and digit string recognition are also furnished in this section. Detailed tabulation of all the test results are furnished in the Appendices [D] to [G].

The Automatic Message and Speaker Recognition (AMSR) system was developed, implemented and tested using Data

General NOVA-840 minicomputer and the whole experimental set-up is as shown in the figure 4.1.

The analog processing part composed of a tape recorder, filter set, A/D converter. Digital processing part was comprised of Nova-840 minicomputer, disk storage system (two drives : one is fixed cartridge, another is removable cartridge type), mag-tape storage system, Tektronix hard copy unit, Tektronix graphics CRT and line printer.

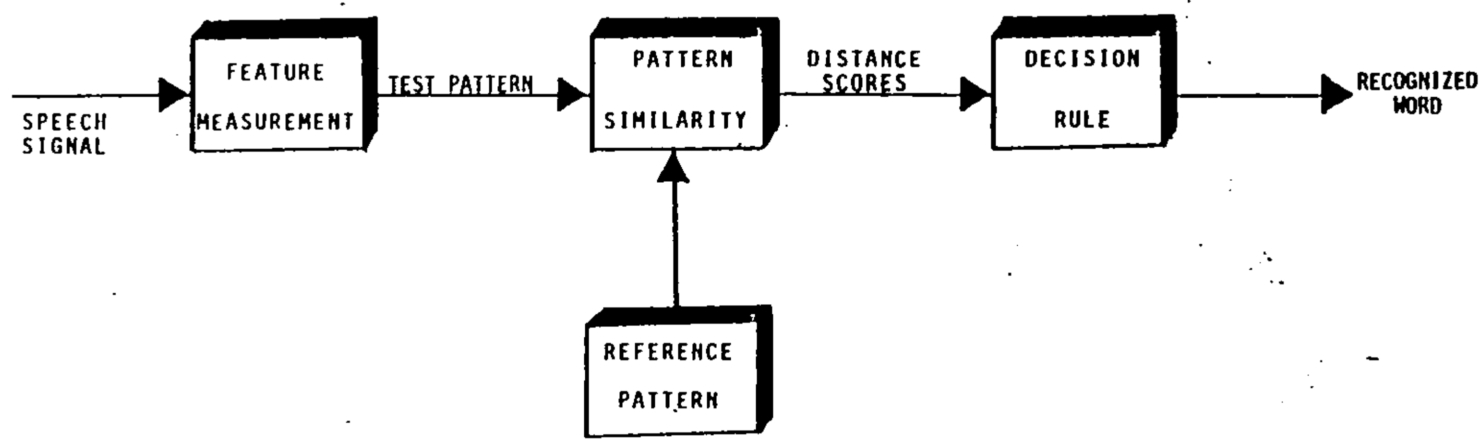


EXPERIMENTAL SET-UP

Figure 4.1

4.2 A SPEAKER-DEPENDENT WORD RECOGNITION SYSTEM

One of the major problems in automatic message recognition systems is the inter-speaker speech differences and size of the vocabulary. An efficient bypass to this problem is to use speaker-dependent automatic message recognition systems, where the system is tuned to the designated speaker. It is for this reason that the initial automatic message recognition systems in practical applications have all used speaker-dependent type of recognition scheme. This type of recognition scheme requires prior training by the speakers for each of the vocabulary entry. During the training phase, reference LPC templates were generated by using the training utterances and word references were stored in the computer library. In the recognition phase any input utterance, after pre-processing and LPC analysis, was compared with reference LPC templates of the designated speaker. The utterance corresponding to the reference LPC pattern which closely resembles the unknown input utterance LPC pattern was the recognized utterance. Figure 4.2 shows the essence of this automatic word recognition scheme.



ESSENCE OF SPEAKER DEPENDENT WORD RECOGNITION

Figure 4.2

Intra-speaker speech variations occur because of differences in speaking rate even though the spoken text is same. These differences in the speaking rates can be minimized by non-linear deformation of the relative time scales of two utterances.

4.2.0.1 DYNAMIC TIME-WARPING

Non-linear deformation of time axis is known as 'Time-Warping' and because dynamic programming approach was utilized for this process, the technique is known as 'Dynamic Time-Warping'.

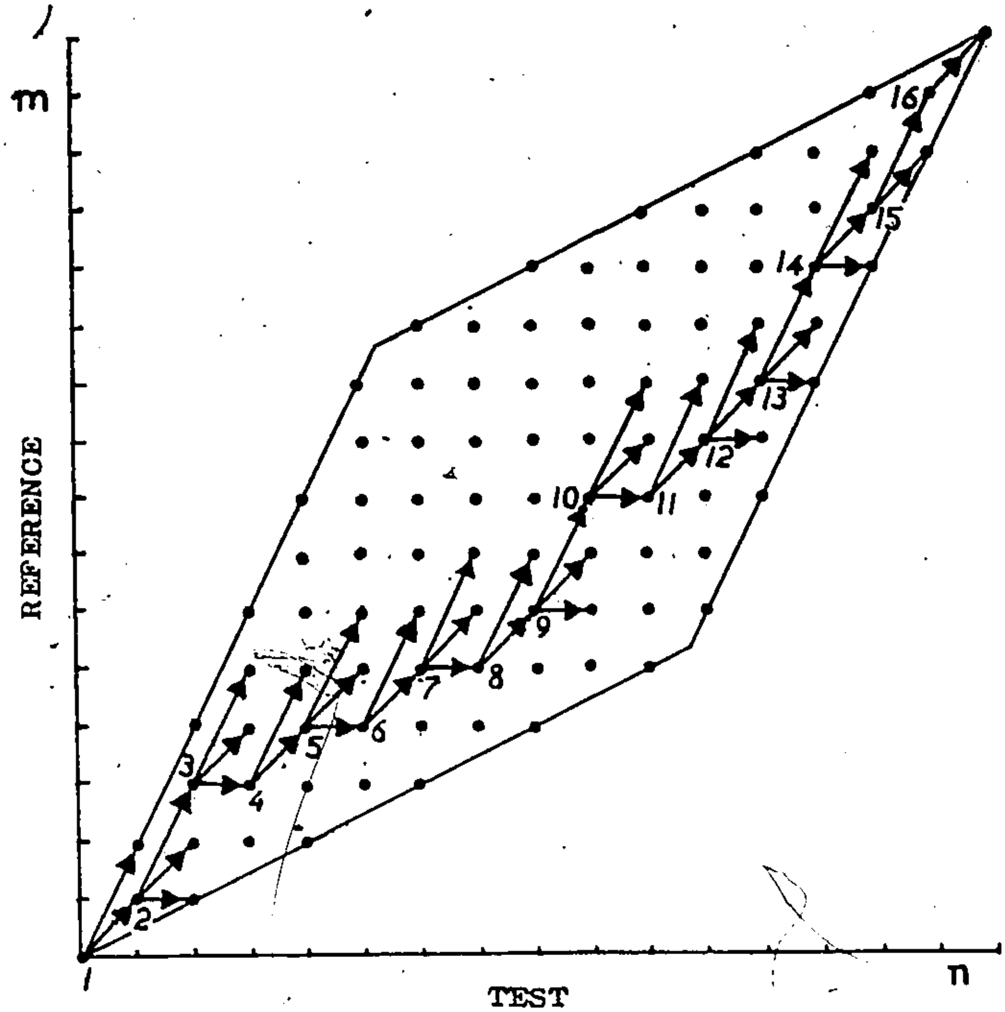
Some sort of normalization is required to facilitate the matching process. Linear time-normalization such as dividing each pattern in time into equal number of subpatterns will not handle local non-linear time variation within the pattern itself. To deal with this problem, a method to match a test pattern against all possible elastic stretchings and shrinkings of each of the reference patterns was proposed by Itakura. A dynamic programming procedure was adopted to perform the following functions :

1. Time alignment of a test LPC pattern against a reference LPC pattern.
2. Multilevel minimizations of local distances between each frame of the test and stretchings, shrinkings of the frames of the reference.

3. Accumulation of minimized local distances at each level of optimization and computation of global distance between the test and a reference pattern.

Figure 4.3 shows 16 level minimization taking place recursively, with simultaneous tracing of minimized warping path and accumulating Itakura's distance values along the path.

Boundary conditions :
$w(1)=1$; $w(16)=16$
Continuity conditions :
$w(n+1)-w(n)=0,1,2$ for $w(n)\neq w(n-1)$
$=1,2$ for $w(n)=w(n-1)$



DYNAMIC PROGRAMMING AND TIME WARPING

Figure 4.3

The details of dynamic programming and time-warping procedure are furnished in the Appendix [C].

4.2.1 PROCESSING FUNCTIONS

4.2.1.1 TRAINING PHASE

All the training utterances were subjected to pre-processing steps, followed by 12 th order LPC extraction in 24 msecs Hamming windowed segments advanced in 50% overlapping steps. LPC templates from second, third, fourth and fifth repetitions were time-warped against LPC pattern of the first utterance. Average LPC pattern from the LPC patterns of first utterance and time-warped LPC patterns of rest of the training utterances was calculated. This represents average reference token for that vocabulary item.

Reference pattern $R(k)$ for each word was stored as a matrix of the form

$$R(k)=[\ c(m;k)\ ,\ b(m;k)\ ]\quad \text{for } m=1,2,\ldots,M(k)\ ;$$
$$k=1,2,\ldots,K$$

where $c(m;k)$ and $b(m;k)$ are the modified parameters of LPC at the m th segment of the k th word reference pattern, $M(k)$ is the number of segments, K is the number of word reference patterns in the library.

Elements of the matrix $R(k)$ were computed using the average LPC template generated in the training phase.

$$c(m;k) = \log(a(m;k)\ a(m;k))$$
$$= \log\Big(\sum_{m=1}^{P} a(m;k)a(m;k)\Big)$$

$c(m;k)$ is the logarithm of the end product of LPC vectors of m th segment of average LPC token for k th word reference pattern (from training phase). $b(m;k)/2$ are the autocorrelation coefficients associated with the inverse filter of the all-pole model and is a vector of the form $(1, b(1), b(2), \ldots, b(12))$ for 12 th order LPC model.

$$b(i) = 2 \sum_{j=0}^{12-i} a(j)\, a(j+i) \;/\; D$$

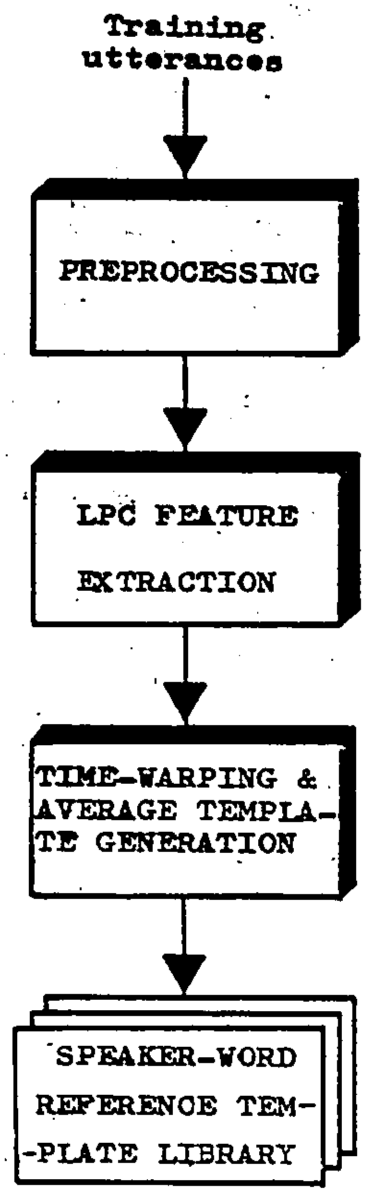
where D is the end product of the LPC vectors of the test utterance.

$$D = (a(i)a(i)) = \sum_{i=1}^{12} a(i)a(i)$$

This procedure of generating a matrix of numbers representing the reference pattern was repeated for each of the vocabulary items(digits 0 to 9) for each of the seven speakers. A data base comprising all speaker-word reference files was generated and stored in the computer disk storage.

An automatic speaker verification system can commit two types of errors. These errors are known as "false rejection error" and "false acceptance error". If a true speaker's identity claim is rejected, the recognition system commits false rejection error. If a false speaker's identity claim is accepted the system makes false acceptance error. Speaker verification thresholds are established such that false rejection and false acceptance errors are equal.

Two test digits from a speaker were time warped against corresponding speaker-digit reference file, thereby forming intra speaker distances. Similarly two test digits (same and different text as in the first case) from all the other speakers, were time warped against the same speaker-digit reference. These distance values form inter speaker distances. Each intra-speaker distance value was taken as the threshold at a time and total of false rejection and false acceptance errors were calculated. One intra-speaker distance value giving equal false rejection and false acceptance errors was established as threshold for that particular input digit. This threshold generation scheme was repeated for all other speaker-digit reference files. Reference tokens containing LPC patterns along with text-independent speaker verification thresholds for all the speaker-digits form reference data base for the experiment. This data was saved in the computer storage. Figure 4.4 shows the processing steps involved in constructing speaker-word reference templates.



SPEAKER-WORD REFERENCE MODEL GENERATION

Figure 4.4

4.2.1.2 RECOGNITION PHASE

Combined speaker verification and digit recognition experiments were conducted using the dynamic time warping of Itakura's distance approach. Test data corresponding to each of the vocabulary entries was used to test the system. A similarity measure known as Itakura's distance was computed in the process of time aligning each frame of a test utterance with different stretching and shrinking of frames of reference tokens.

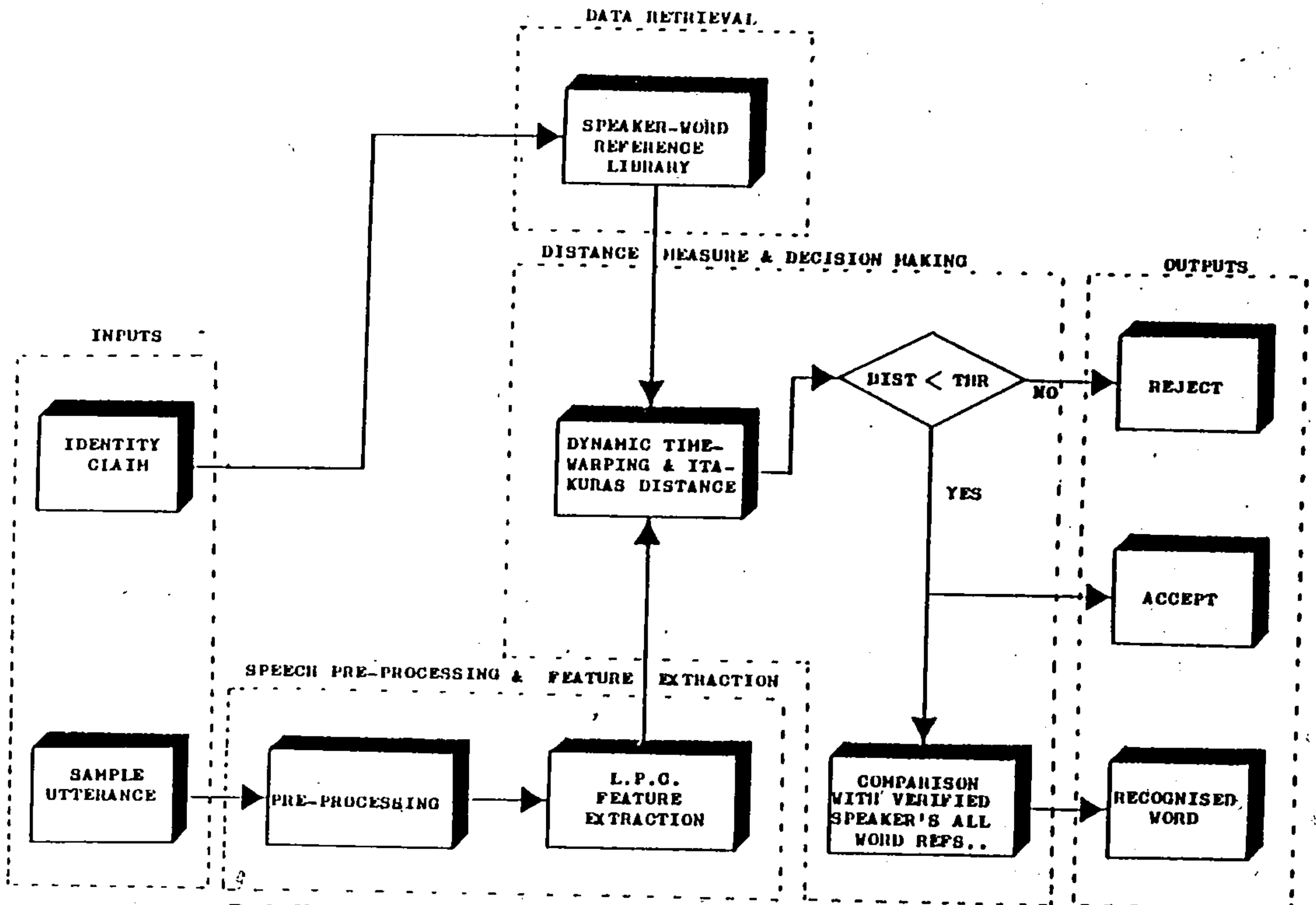
$$d(n,m;k) = c(m;k) + \log[(b(m;k)r(n)) / (a'(n)r(n))]$$

where $d(n,m;k)$ is the Itakura's distance between the n th segment of the test and m th segment of the k th reference pattern; $c(m;k)$ and $b(m;k)$ are components of the reference token matrix; $a'(n)$ is the LPC vector of the form $(1,a(0),a(1),\ldots,a(12))$ of the n th segment of the test; $r(n)$ is the autocorrelation coefficients vector of the form $(1,r(1),\ldots r(12))$ for the n th segment of the test, where $r(n)$ coefficients are calculated as

$$r(m) = (1/N) \sum_{i=1}^{N-m} s(i)s(i+m) \qquad \text{for } m=0,1,\ldots,12$$

N is the total number samples in the Hamming windowed segment. $(a'(n)r(n))$ represents the end product of two vectors $a'(n)$ and $r(n)$ of the n th segment of the test digit.

A block diagram of all the experimental procedures conducted in connection with simultaneous verification of speaker and recognition of the spoken word are as shown in the figure 4.5. Identity claim of a speaker and a sample utterance form inputs to the system. Using this identity claim, the algorithm retrieves claimed speaker's all word reference tokens and corresponding thresholds from the reference data base. At the same time the test utterance undergoes pre-processing and 12 th order LPC analysis in 24 msecs Hamming windowed segments advanced in 50% overlapped steps. The test utterance LPC pattern was aligned against each of the reference tokens of the designated speaker using dynamic time-warping procedure. If the distance value lies below the threshold the speaker's identity claim was accepted, otherwise the identity claim was rejected. If the identity claim was accepted, then the claimed speaker's utterance corresponding to a reference token giving minimum distance from the test LPC pattern was taken as the recognized word. If the speaker's identity claim was rejected no further processing was carried out for word recognition.



BLOCK DIAGRAM OF COMBINED SPEAKER AND DIGIT RECOGNITION

Figure 4.5

Detailed tabulation of results of combined text independent speaker verification and speaker dependent digit recognition are presented in Appendix [E]. Overall total errors (false rejection and false acceptance) and overall recognition accuracies are given in Table 4.1.

Total number of tests for text-independent speaker verification = 6120

Total number of tests for speaker-dependent digit recognition = 1200

SPEAKER RECOGNITION AND DIGIT RECOGNITION
(LPC analysis with Itakura's similarity measure)

	Text-independent speaker verification	speaker-dependent digit recognition
Total errors	799	64
% Overall recognition accuracy	86.94	94.66

Table 4.1

4.3 SPEAKER AND DIGIT RECOGNITION USING ORTHOGONAL PARAMETERS

It is well known that we can identify a person from the sound of his voice, yet we frequently observe that two different voices sound alike. The variation in voices has made automatic message recognition difficult, while the similarities have limited the success of automatic speaker recognition.

The general approach is to extract some acoustic attributes from one's speech and compare them with a reference set previously stored in the machine's library. If there is a close resemblance between the test and reference features, the speaker is said to be recognized. In 'text-dependent speaker recognition' where the test and reference features are obtained from the same text material. However in the 'text-independent' case, the test and reference text bear no linguistic relationship to each other. It is because these acoustic attributes derived from speech, not only signify the inter-speaker variations, but also are functions of the speech text. Hence the success of a 'text-independent automatic speaker recognition system' depends on the extraction of a set of acoustic properties that can characterize each speaker, independent of the speech text.

Earlier research conducted by Sambur [5], and Robert E. Bogner [15], applied the method of orthogonal measurements which were sets of LPCs, reflection coefficients, and logarithmic area coefficients. They proposed that highly accurate speaker verification could be achieved, independent of speech transmission medium and spoken text. Research conducted here concentrates on the problem of combined speaker verification and digit recognition using orthogonal LPCs, Inverse Filter Spectral Coefficients (IFSCs), Direct Fourier Transform of speech (DFTs). Also in this work subsets of orthogonal LPCs were examined for their speaker sensitive and speech sensitive characteristics.

To find the resemblance between test and reference templates of orthogonal parameters, two similarity measures were utilized. These similarity measures are known as 'distance of averages' and 'average distance'. The computational details of these measures are given below.

DISTANCE OF AVERAGES :

$$d1 = \sum_{\substack{i \\ \text{chosen subset} \\ \text{of orth.pars.}}} [\, \phi_{im} - z_{im}]^2 / \lambda_{im}$$

where ϕ_{im} is i th overall average orthogonal parameter for the reference utterance of m th speaker

z_{im} is the i th overall average orthogonal parameter for the test utterance of m th speaker.

λ_{im} is the i th eigen value of the reference utterance of m th speaker.

This similarity measure is computed between overall average orthogonal parameters of the test and reference utterances. Overall average orthogonal parameters of the reference is derived from training utterances. Element-wise sum of orthogonal parameters is generated across all the frames of all the training utterances. Average of these individual element-wise sums, over all the frames give overall average orthogonal parameters. This similarity measure essentially brings out the resemblance between the test and reference utterances in global or overall average space, rather than frame-wise space. Non-linear time warping is not meaningful, while utilizing this similarity measure.

AVERAGE DISTANCE:

$$d2 = (1/N)\sum_{i=1}^{N}\sum_{\substack{k \\ \text{chosen subset} \\ \text{of orth.pars.}}}[\phi_{ik}\text{test} - \phi_{ik}\text{ref}]^2/\lambda_k\text{ref}$$

where N is the total number of frames in test or reference utterance.

ϕ_{ik} test is the k th othogonal parameter of i th frame of the test utterance.

ϕ_{ik} ref is the k th orthogonal parameter of i th frame of the reference utterance.

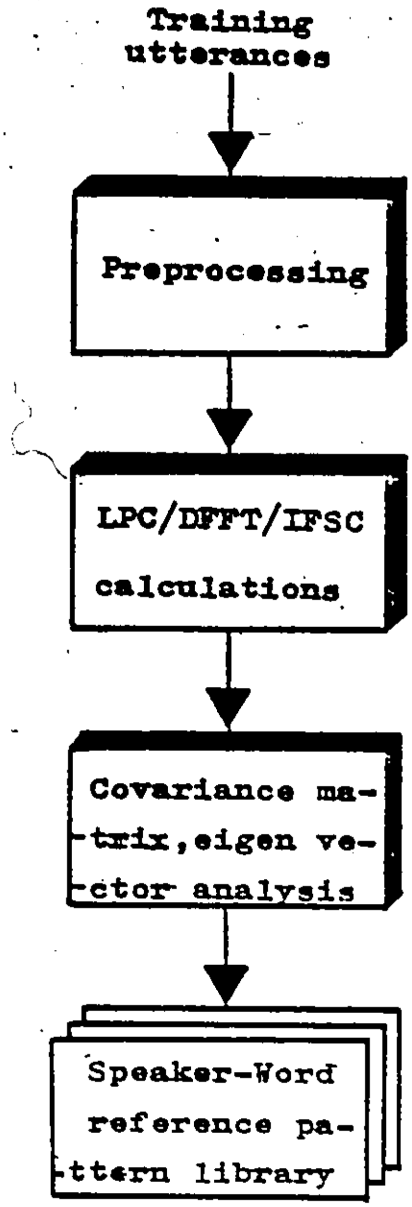
λ_k ref is k th eigen value of the reference utterance.

This similarity measure is computed between framewise orthogonal parameters of the test and reference utterances. Frame-wise orthogonal parameters of the reference were derived from the training utterances. Average frame-wise orthogonal parameters of all the training utterances, form reference pattern of orthogonal parameters. This similarity measure essentially brings out the resemblance between orthogonal parameters of individual frames of test and reference utterances. Average of these frame-wise orthogonal parameter distances formulate as 'average distance'. Non-linear time warping of test against reference utterances is meaningful, while utilizing this similarity measure.

The aim of the experiment was to find an efficient scheme for simultaneous speaker verification and message recognition. Speech data for the experiment comprised of spoken digits 0 to 9 collected from seven speakers in different sessions. This experiment proceeds in two phases viz : training phase and recognition phase.

4.3.1 TRAINING PHASE

Five repetitions of each of spoken digit from each of the speaker was used to train the system. Total number of Hamming windowed segments in an utterance was made equal to 30 by changing the length of overlap between adjacent segments. LPC, DFT and IFSC analysis was performed on these utterances. Covariance matrix of the individual features from five training utterances, was generated. Eigen vectors of this covariance matrix were calculated. Reference orthogonal parameters were generated by utilizing these reference eigenvectors. In order to compute two similarity measures 'average distance' and 'distance of averages', overall average and frame-wise overall average orthogonal parameters were calculated. An overall average feature vector was calculated by summing individual elements of the vector over all the frames and over all the training utterances. This was followed by calculation of the average by dividing the overall total by the total number of frames in all the training utterances. An overall average orthogonal feature vector was generated by evaluating a dot product of the overall average feature vector and reference eigen vector. Frame-wise overall average feature vectors were generated by summing frame-wise feature vectors with corresponding frame-wise feature vectors of all other training utterances. Frame-wise overall average orthogonal parameters were generated by calculating the dot product of reference eigenvectors and frame-wise overall average feature vectors. Overall average orthogonal features and overall frame-wise average orthogonal parameters were stored in the computer library. Figure 4.6 shows the different processing functions taking place in the training phase.



<u>SPEAKER-WORD REFERENCE MODEL GENERATION</u>

Figure 4.6

Thresholds for text-dependent and text-independent speaker verification experiments were generated. Separate sets of thresholds were generated in each of the feature spaces viz : orthogonal LPCs, IFSCs, DFTs.

THRESHOLDS FOR TEXT-DEPENDENT SPEAKER VERIFICATION :

Five test digits from a speaker were compared with the corresponding speaker-digit reference file. These distance values form intra-speaker distances. Similarly five test digits (same text as in the first case) from all the speakers, were compared with the same speaker-digit reference. These distance values form inter-speaker distances. Each of the intra-speaker distance was taken as threshold at a time and total false rejection and false acceptance errors were calculated. One intra-speaker distance giving equal false rejection error and false acceptance error was established as the verification threshold. Two exclusive thresholds per speaker-digit were generated for two similarity measures. This procedure of establishing text-dependent speaker verification thresholds was repeated with all the speaker-digit references.

THRESHOLDS FOR TEXT-INDEPENDENT SPEAKER VERIFICATION :

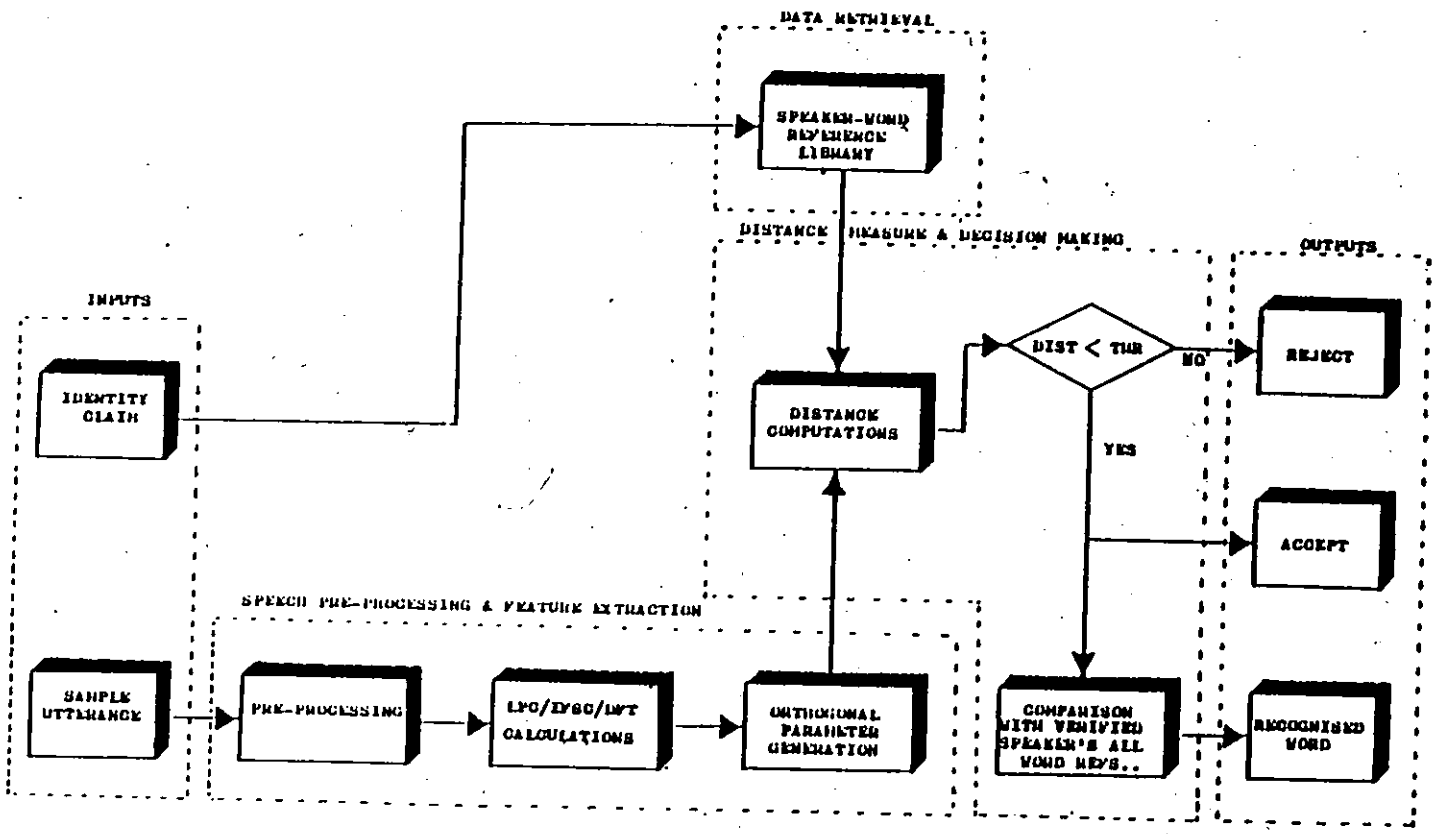
Five test digits from a speaker, were compared with corresponding speaker-digit reference file, thereby forming intra-speaker distances. Similarly five test digits (same and different text as in the first case) from all the other speakers, were compared with the same speaker-digit reference. These distance values form inter-speaker distances. Each of the intra-speaker distance was taken as a threshold at a time and total false rejection and false acceptance errors were computed. One intra-speaker distance giving equal false rejection and false acceptance errors was established as verification threshold. Two exclusive thresholds per speaker-digit were generated compatible for average distance and distance of averages. This procedure of computation of text-independent speaker verification thresholds was repeated for all the speaker-digit reference files.

The following patterns were stored in the computer library as reference data base.

1. Text-dependent and text-independent speaker verification thresholds.
2. Overall average orthogonal parameters (for LPCs, IFSCs, DFTs) for all speaker-digit utterances.
3. Overall frame-wise average orthogonal parameters (LPCs, IFSCs, DFTs) for all speaker-digit utterances.

4.3.2 RECOGNITION PHASE

Figure 4.7 shows different processing steps involved in the recognition phase. Identity claim and the sample utterance comprising the test digit, form inputs to the system. Text-independent speaker verification and simultaneous digit recognition was conducted. Identity claim was used to retrieve the claimed speaker's digit reference files and corresponding text-independent verification thresholds. Orthogonal parameters (LPCs, IFSCs, DFTs) were derived from the sample test utterance. Two similarity measures were computed to find the resemblance between the sample utterance and claimed speaker's reference files. If the computed distance lies below the text-independent threshold, the speaker's identity claim was accepted. If the computed distance lies above any of the thresholds, the speaker's identity claim was rejected and no further operation was done to recognize the spoken digit. However if the speaker's identity claim was accepted, the algorithm proceeds to recognize the spoken digit. The distances between the sample utterance and the claimed speaker's all digit files were computed. The digit utterance corresponding to a reference file giving minimum distance with the sample test data was reported as the recognized digit.



BLOCK DIAGRAM OF COMBINED SPEAKER AND DIGIT RECOGNITION

Figure 4.7

Five repetitions of test digits (these repetitions were different from those used for training) from all the speakers, were used to test the system. In the text-independent mode of the experiment five repetitions of all the speakers of all the digits were input to the system. Whereas in the case of text-dependent mode of the experiment, same digit repetitions from all the speakers were used to test the system. Total errors (false acceptance + false rejection) committed by the system and the recognition accuracy (100 - % errors) were calculated. This procedure of calculating total errors and recognition accuracy was repeated with all speaker-digit references. The numerical tabulation of text dependent speaker verification results are presented in Appendix [D]. Detailed numerical tabulations of combined text independent speaker verification and speaker dependent digit recognition results are furnished in the Appendix [E]. Overall total errors and overall recognition accuracies for combined speaker and digit recognition systems operating with orthogonal parameters (LPCs, IFSCs, DFTs, LPC subsets) with two distance measures are presented in Table 4.2.

RECOGNITION OF SPEAKER AND SINGLE DIGIT UTTERANCES

Total number of tests for text-independent speaker verification = 21350

Total number of tests for speaker-dependent digit recognition = 3500

OVERALL RECOGNITION ACCURACIES (IN %)
(similarity measure : average distance)

PARAMETER SETS		TEXT-INDEPENDENT SPEAKER VERIFICATION	SPEAKER-DEPENDENT DIGIT RECOGNITION
LPCs	subset 1-8	98.47	92.06
	subset 4-9	98.12	90.14
	subset 9-12	97.94	90.40
	set 1-12	98.98	94.32
IFSCs	set 1-32	97.52	93.49
DFTs	set 1-32	99.46	97.83

OVERALL RECOGNITION ACCURACIES (IN %)
(similarity measure : distance of averages)

PARAMETER SETS		TEXT-INDEPENDENT SPEAKER VERIFICATION	SPEAKER-DEPENDENT DIGIT RECOGNITION
LPCs	subset 1-8	97.28	75.80
	subset 4-9	97.82	85.91
	subset 9-12	98.11	88.66
	set 1-12	98.66	93.91
IFSCs	set 1-32	97.31	94.97
DFTs	set 1-32	97.55	89.88

Table 4.2

4.3.2.1 DISCUSSION OF RESULTS

Simultaneous speaker and digit recognition potential based on single digit utterances is reviewed here from different perspectives.

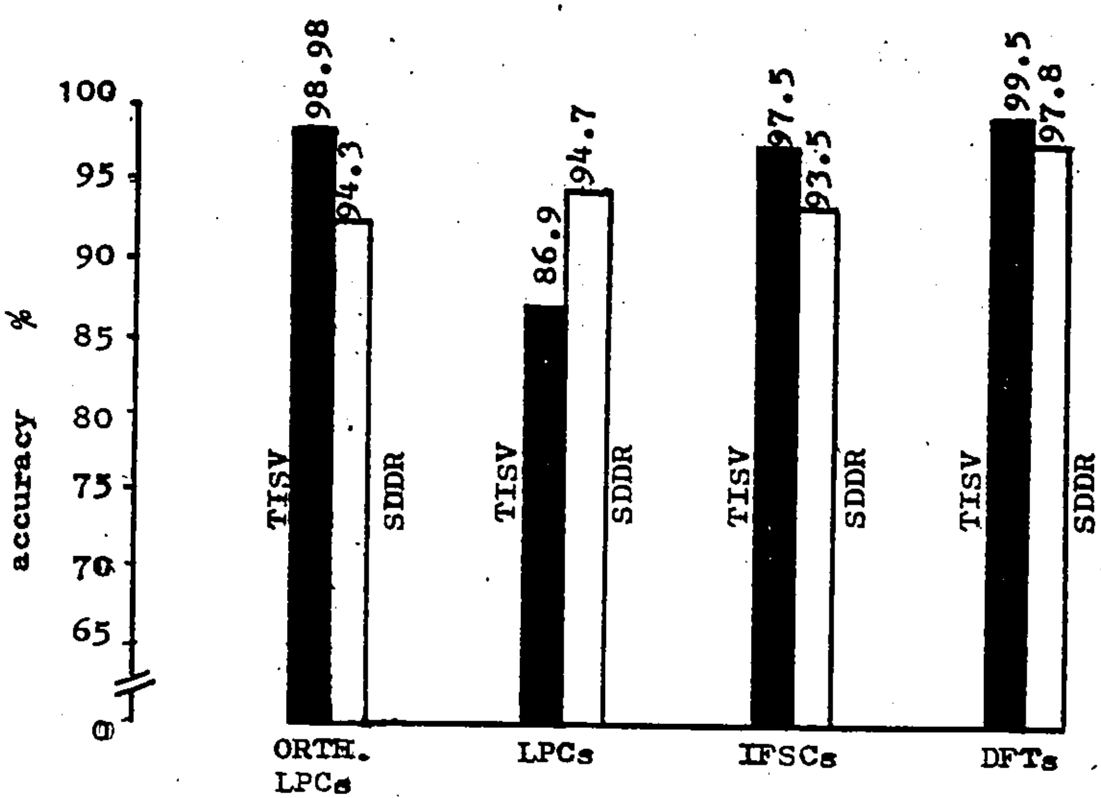
1. LPCs with dynamic time warping, Itakura's distance was basically a message recognition approach. The experiments investigating simultaneous speaker and digit recognition, gave 86.94% speaker verification accuracy and 94.66% digit recognition accuracy. This strengthens the suggestion that this method of recognition is more suitable for message recognition than speaker recognition. This can be attributed to the fact that dynamic time warping is an efficient tool for non-linear template matching, thereby bringing out linguistic similarities better than speaker discrimination characteristics.
2. LPCs are known to be mutually correlated. An orthogonalization step is added to make these parameters mutually uncorrelated, thereby bringing out speaker and message sensitive characteristics. The experimental investigations with orthogonal LPCs gave 12.04% better speaker verification accuracies than LPCs without orthogonalization. However LPCs (with Itakura's distance) gave 0.34% better digit recognition accuracies than orthogonal LPCs.

3. Itakura's distance in LPC space, average distance and distance of averages in orthogonal parameter space are compared. Average distance is the best candidate for simultaneous speaker and digit recognition viewpoint. Average distance giving frame-wise similarity scores in orthogonal parameter space, is a strong and attractive feature as compared to other similarity measures. Itakura's distance also generates frame-wise distance by non-linearly time warping the utterances, but in LPC space.
4. A comparison of performance of simultaneous speaker and digit recognition systems operating with orthogonal LPCs, IFSCs, DFTs, LPCs with Itakura's distance, is made. Orthogonal DFTs are the best candidate features for simultaneous speaker and digit recognition. This can be attributed to the fact that DFTs are orthogonal spectral parameters derived directly from speech. Bar chart shown in the figure 4.8 depicts the performance of different feature sets in terms of text-independent speaker verification and speaker dependent digit recognition accuracies.
5. Review of results obtained with combined recognition schemes operating with orthogonal LPC sets 1 to 12, 9 to 12, 1 to 8, 4 to 9 indicates that set 1 to 12 is the best set from simultaneous speaker and digit recognition viewpoint. This reveals the fact that

accuracies improve with increasing number of elements in a set, for distance computations for single digit utterance case. Bar charts shown in figure 4.9 show the analysis results obtained with different subsets of orthogonal LPC features, with single digit utterances. Performances of average distance and distance of averages can also be seen from these bar charts.

6. Speaker independent digit recognition using orthogonal parameters is not feasible. Hence initially speaker's identity claim is to be verified, then, speaker dependent digit recognition is to be conducted. Combined recognition of speaker and digits was accomplished in two stages in the algorithm viz : verification of the identity claim of the speaker in text independent mode, followed by claimed speaker's digit recognition in speaker dependent mode.

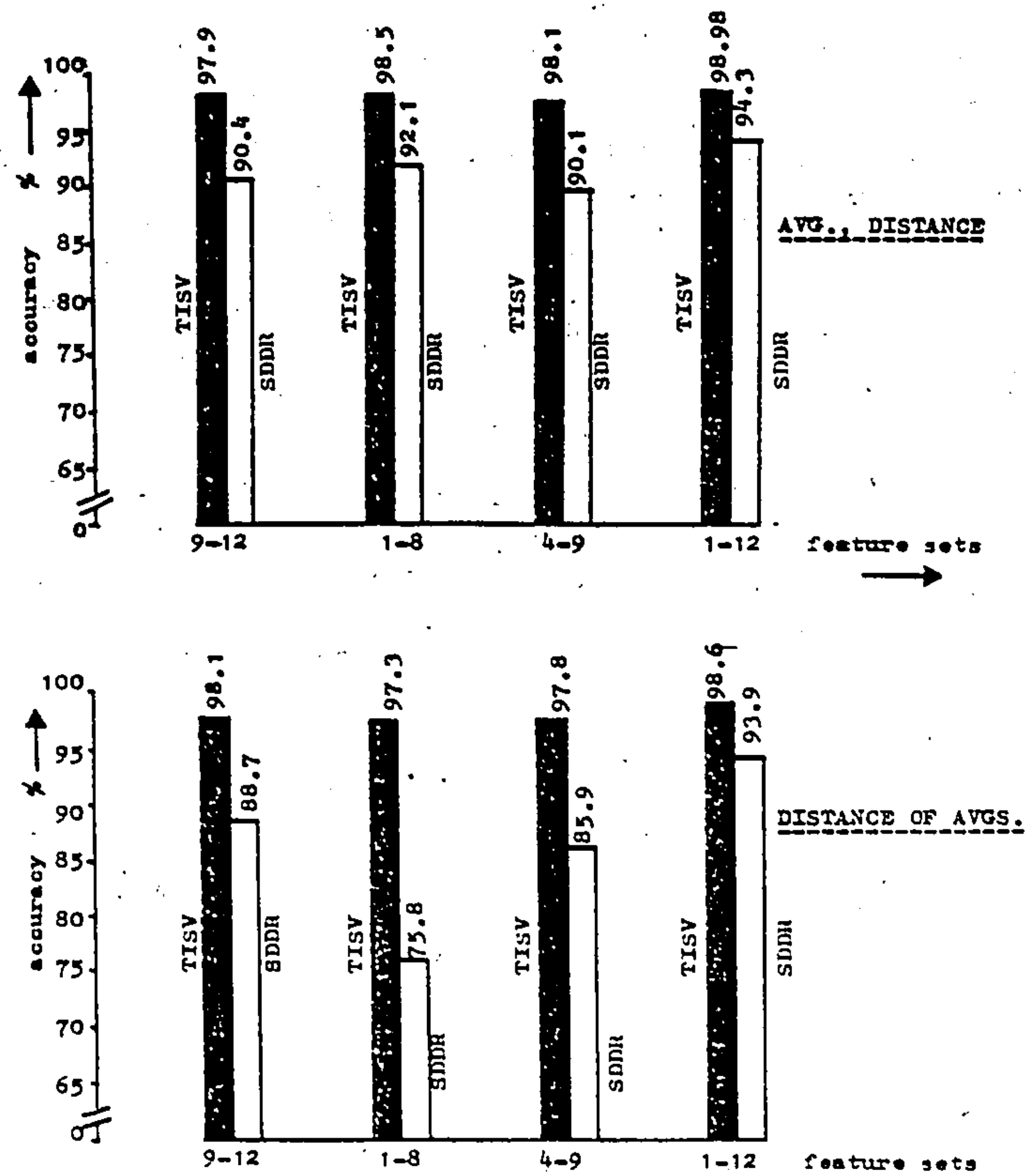
7. Pooling all the best candidates formulates the best scheme of simultaneous speaker and digit recognition. Orthogonal DFTs with average distance gave 99.46% speaker verification and 97.83% digit recognition accuracies.



TISV : Text Independent Speaker Verification
SDDR : Speaker Dependent Digit Recognition

PERFORMANCE EVALUATION OF DIFFERENT FEATURES FOR SIMULTANEOUS SPEAKER & DIGIT RECOGNITION

Figure 4.8



TISV : Text Independent Speaker Verification
SDDR : Speaker Dependent Digit Recognition

LPC SUBSET PERFORMANCE FOR SIMULTANEOUS SPEAKER & WORD RECOGNITION USING SINGLE DIGIT UTTERANCES

Figure 4.9

4.3.3 DIGIT STRINGS FOR COMBINED MESSAGE & SPEAKER RECOGNITION

The need for enhancing the recognition accuracies led to the investigation of 3 digit sequences and 7 digit sequences as input speech data for combined recognition of speaker and spoken digits. Vocabulary for 3 digit strings composed of '387', '210', '777', '888', '213', '877', '037' and that for 7 digit strings composed of '2536879' and '3689427' respectively. Orthogonal parameter (LPCs, DFTs) analysis was conducted on these digit string utterances. The feasibility of the proposed algorithm for single digit utterances was investigated for simultaneous speaker verification and digit string recognition.

Here the 3 digit strings or 7 digit strings were taken as test and reference blocks of data for recognition. Reference files of orthogonal LPCs,orthogonal DFTs were generated as explained in the single digit case,only difference being that instead of single digits, digit strings were used for each repetition. These digit string repetitions were used to construct covariance matrix and eigen vectors. Test utterances (not used in the training phase) were used to construct test digit strings. Test digit strings were used to calculate text-dependent and text-independent speaker verification thresholds. Text-dependent, text-independent thresholds, overall average orthogonal parameters, overall frame-wise average orthogonal parameters were stored as reference data base in the computer. Experiments for combined text-independent speaker verification and speaker-dependent digit string recognition were conducted. Detailed numerical tabulation of results of combined speaker and 3 digit string recognition experiments are presented in Appendix [F]. Detailed numerical tabulation of results of combined speaker and 7 digit string recognition experiments are furnished in Appendix [G]. Total errors (false acceptance + false rejection errors) and overall recognition accuracies for 3 digit and 7 digit string utterances are presented in Table 4.3 and Table 4.4 respectively.

RECOGNITION OF SPEAKER AND 3-DIGIT STRINGS

Total number of tests for text-independent speaker verification	10535	(for LPCs)
	1995	(for DFTs)
Total number of tests for speaker-dependent digit string recognition	1715	(for LPCs)
	315	(for DFTs)

OVERALL RECOGNITION ACCURACIES (IN %)
(similarity measure : average distance)

PARAMETER SETS		TEXT-INDEPENDENT SPEAKER VERIFICATION	SPEAKER-DEPENDENT DIGIT RECOGNITION
LPCs	subset 1-8	99.40	95.30
	subset 4-9	98.40	87.10
	subset 9-12	97.90	76.20
	set 1-12	99.30	91.30
DFTs	set 1-32	99.60	100.00

OVERALL RECOGNITION ACCURACIES (IN %)
(similarity measure : distance of averages)

PARAMETER SETS		TEXT-INDEPENDENT SPEAKER VERIFICATION	SPEAKER-DEPENDENT DIGIT RECOGNITION
LPCs	subset 1-8	98.70	83.20
	subset 4-9	98.10	77.70
	subset 9-12	97.40	71.40
	set 1-12	99.40	87.20
DFTs	set 1-32	97.30	98.20

Table 4.3

RECOGNITION OF SPEAKER AND 7-DIGIT STRINGS

Total number of tests for text-independent speaker verification 910 (for LPCs) 450 (for DFTs)

Total number of tests for speaker-dependent digit string recognition 140 (for LPCs) 100 (for DFTs)

OVERALL RECOGNITION ACCURACIES (IN %)
(similarity measure : average distance)

PARAMETER SETS		TEXT-INDEPENDENT SPEAKER VERIFICATION	SPEAKER-DEPENDENT DIGIT RECOGNITION
LPCs	subset 1-8	100.00	100.00
	subset 4-9	100.00	97.86
	subset 9-12	92.53	88.57
	set 1-12	98.68	95.00
DFTs	set 1-32	100.00	100.00

OVERALL RECOGNITION ACCURACIES (IN %)
(similarity measure : distance of averages)

PARAMETER SETS		TEXT-INDEPENDENT SPEAKER VERIFICATION	SPEAKER-DEPENDENT DIGIT RECOGNITION
LPCs	subset 1-8	98.57	73.57
	subset 4-9	95.27	64.29
	subset 9-12	90.00	60.71
	set 1-12	95.93	66.43
DFTs	set 1-32	99.11	70.00

Table 4.4

4.3.3.1 DISCUSSION OF RESULTS

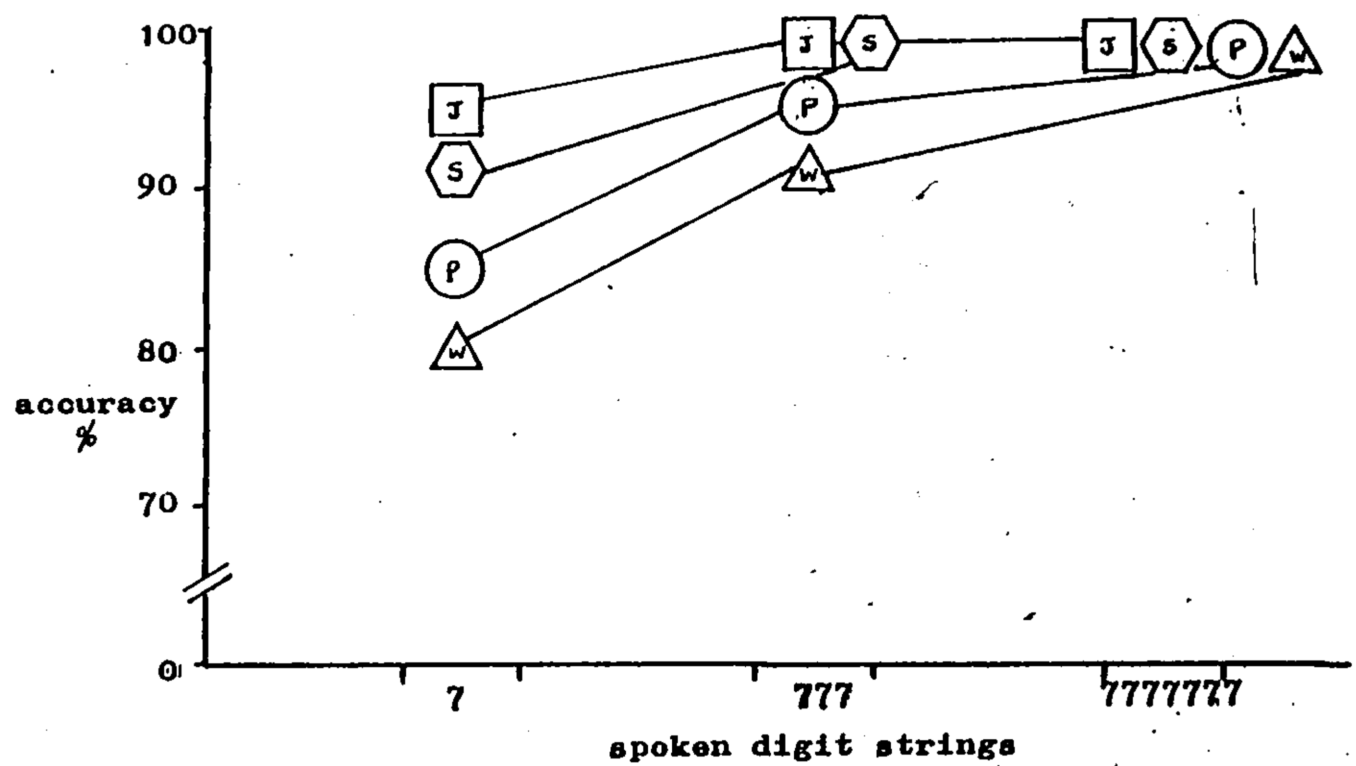
These experimental investigations on spoken digit strings may be treated as preliminary, as the speech data base (digit strings 387, 210, 777, 888, 213, 877, 037, 2536879, 3689427) was not statistically large enough to establish generalized conclusions. Computer disk space and memory space problems for accomodating larger number of digit strings restricted the investigation to only limited digit sequence vocabulary.

Simultaneous speaker verification and digit string recognition potential by different methods is reviewed here.

1. Performance of combined recognition systems operating with average distance and distance of averages is compared. Average distance gave the best performance from simultaneous speaker and digit recognition accuracy viewpoints.
2. Feature-wise review of performance indicate that both LPCs and DFTs (in orthogonal parameter space) gave identical speaker verification accuracies, for 3 digit and 7 digit string utterances.
3. Performance evaluation of orthogonal LPC parameter subsets is conducted. Both sets 1 to 12 and 1 to 8 gave identical speaker verification performance (99.3% and 99.4% respectively) for 3 digit string utterances. In the case of 7 digit string utterances, orthogonal LPC subset 1 to 8 gave the best

performance (100% text independent speaker verification) as compared to other sets and subsets. This reveals the fact that least significant eight orthogonal LPCs are adequate to represent speaker characteristics, especially in the case of utterances longer than single digits.

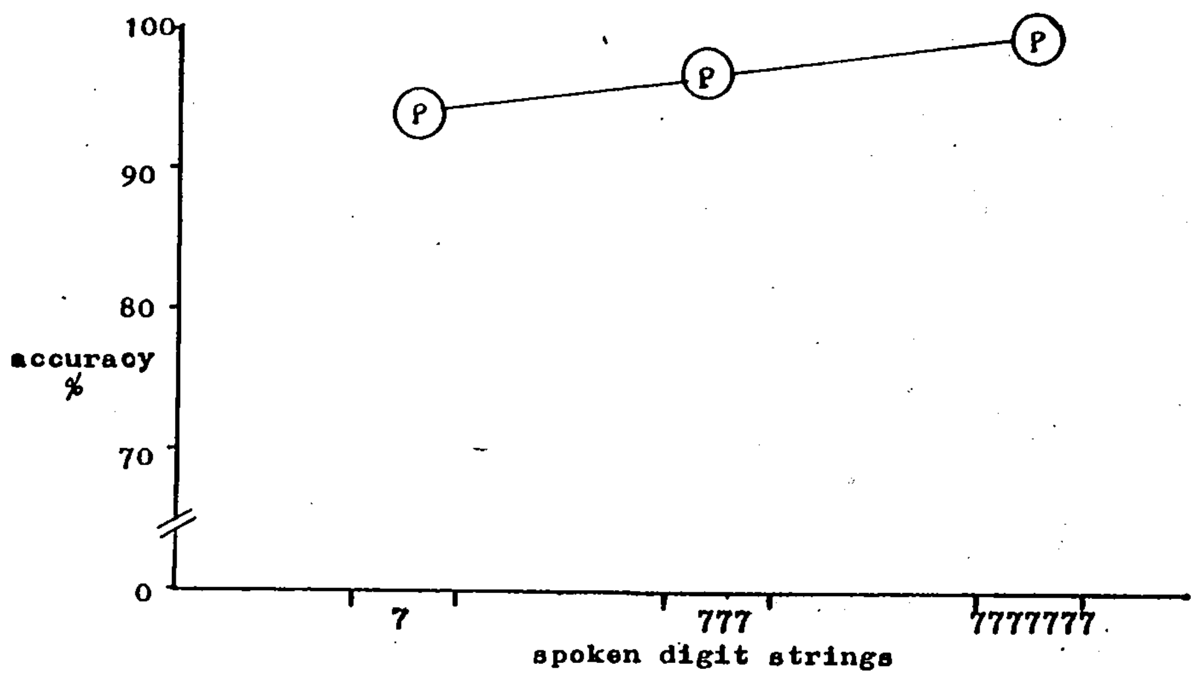
4. Experimental results reveal that machine's capability to recognize some speakers and their spoken digits, was uniformly better than that of other speaker-digits. Speaker-wise perspective of improvement of text dependent, text independent speaker verification and speaker dependent digit/digit string recognition accuracies are shown in the graphs in figures 4.10, 4.11 and 4.12 respectively. The letters at discrete points indicate the speakers. The remaining speakers not appearing in the graphs, uniformly gave 100% accuracies.



(L.P.C. ANALYSIS)

IMPROVEMENT IN TEXT-DEPENDENT SPEAKER VERIFICATION
WITH DIGIT SEQUENCE INPUTS
(SPEAKER-WISE PERSPECTIVE)

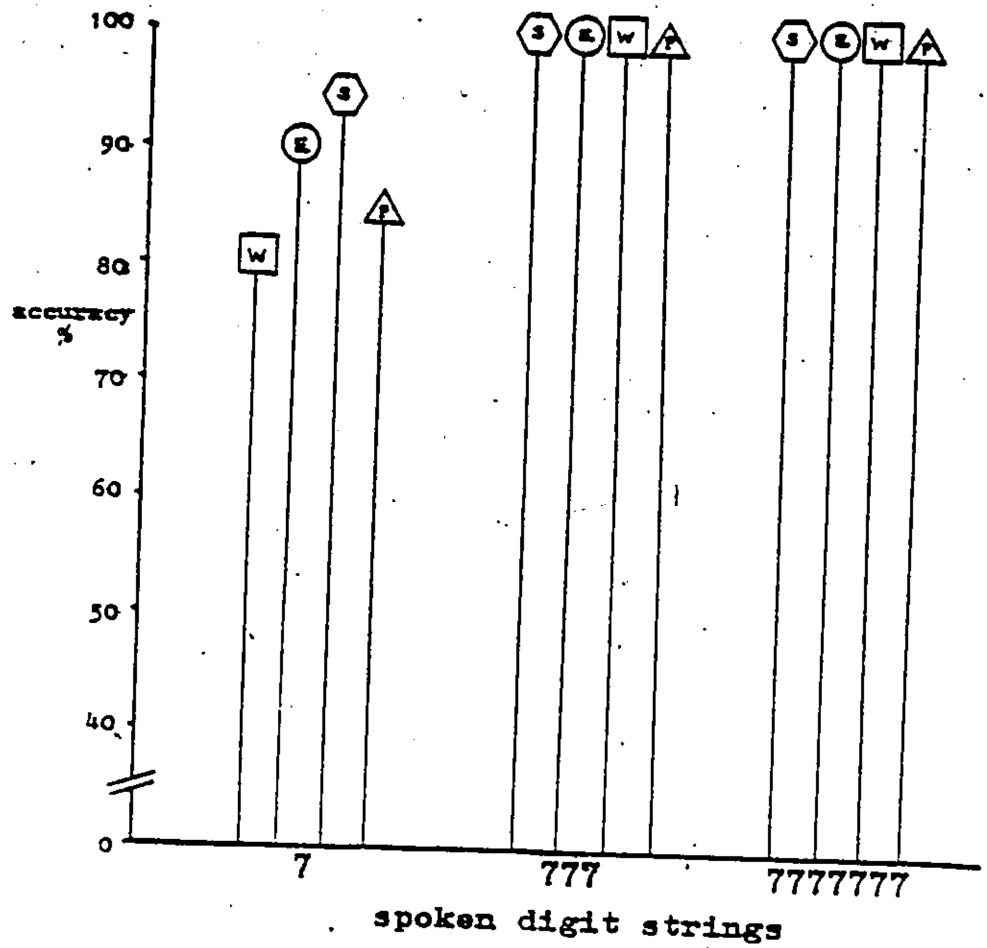
Figure 4.10



(L.P.C. ANALYSIS)

IMPROVEMENT IN TEXT-INDEPENDENT SPEAKER VERIFICATION
WITH DIGIT SEQUENCE INPUTS
(SPEAKER-WISE PERSPECTIVE)

Figure 4.11



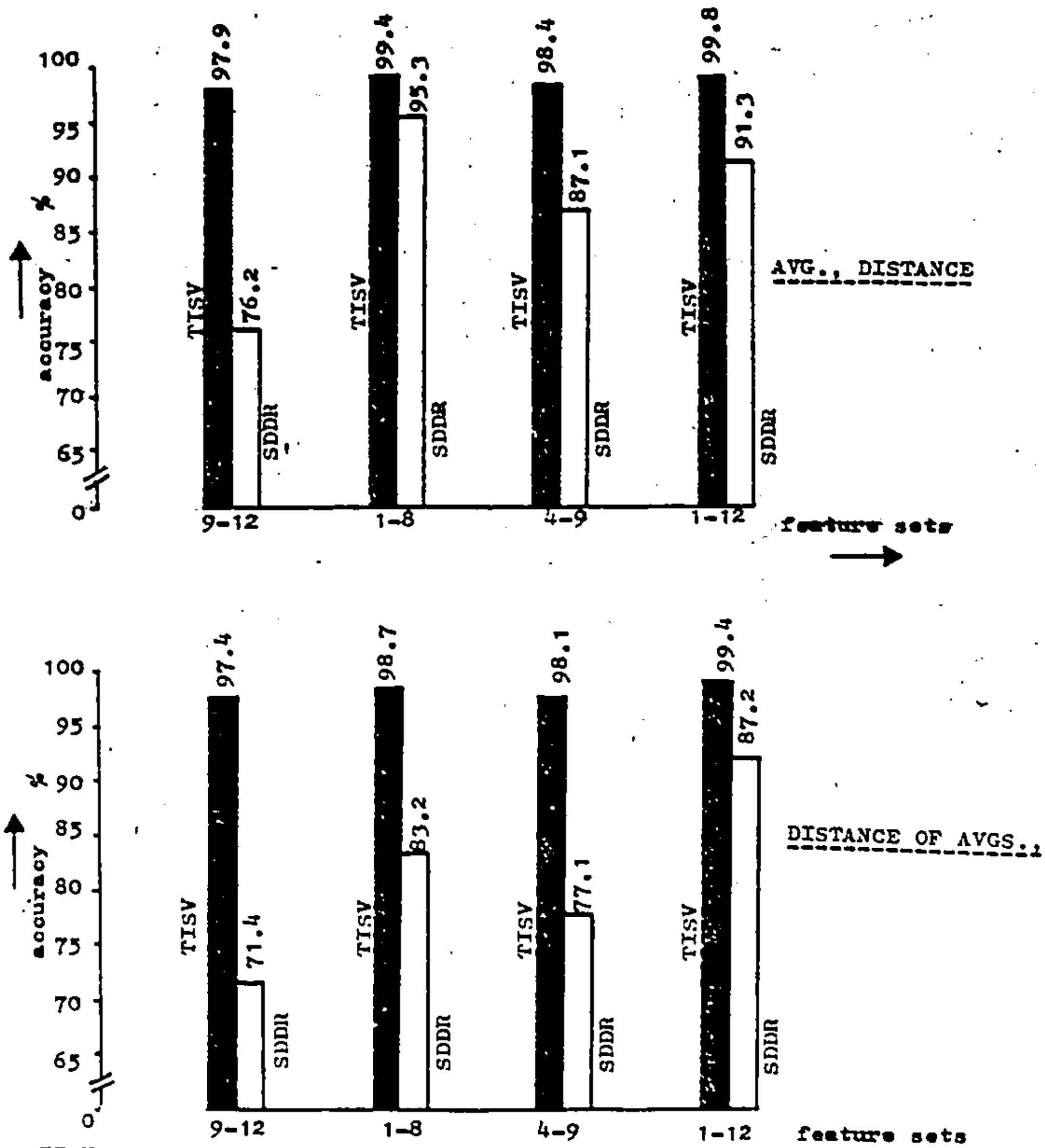
(L.P.C. ANALYSIS)

IMPROVEMENT IN DIGIT/DIGIT STRING RECOGNITION ACCURACY WITH DIGIT SEQUENCE INPUTS

(SPEAKER-WISE PERSPECTIVE)

Figure 4.12

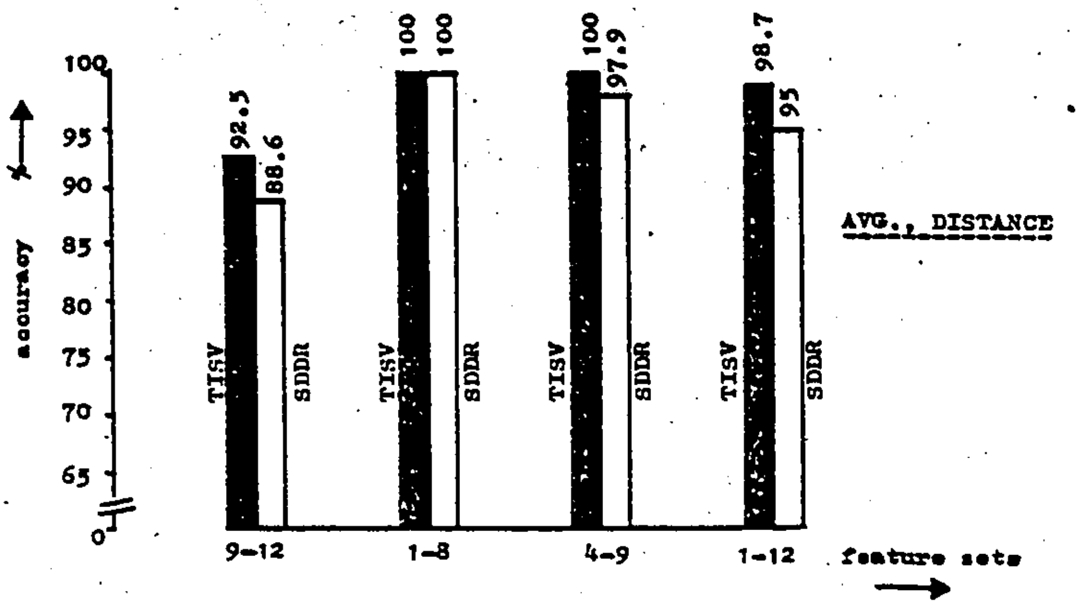
Performance of different subsets of orthogonal LPCs for combined speaker and digit string recognition are depicted in figures 4.13 and 4.14 in the form of bar charts. Figure 4.13 shows the performance of average distance and distance of averages with different subsets of orthogonal LPCs for 3 digit string utterances. Figure 4.14 depicts the performance of average distance and distance of averages with different subsets of orthogonal LPCs for 7 digit string utterances. Figure 4.15 shows the effect of digit strings on text independent speaker verification with orthogonal parameters (LPCs, DFTs).

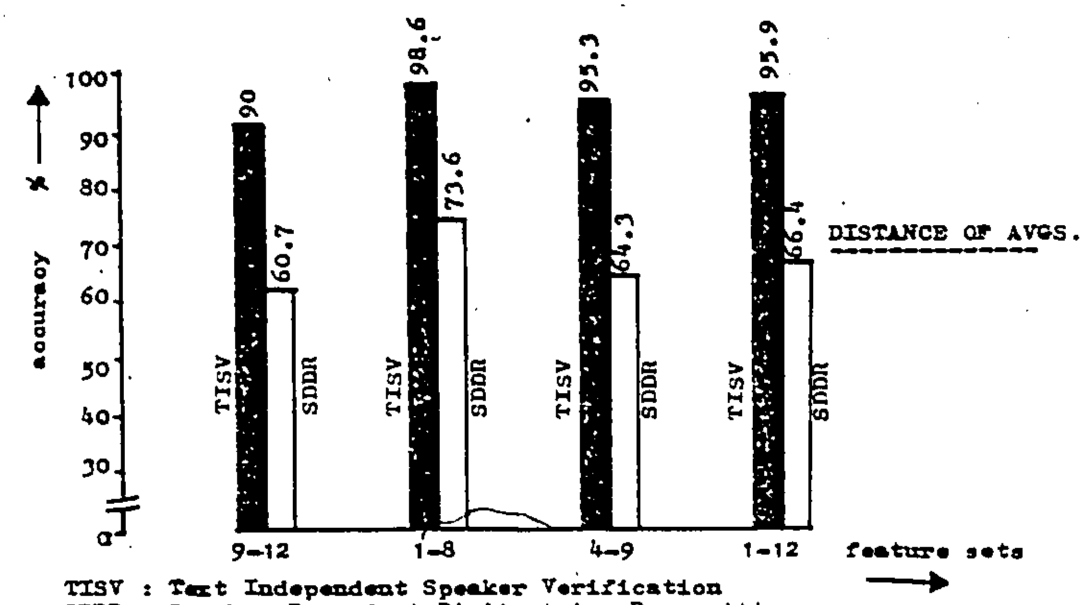


TISV : Text Independent Speaker Verification
SDDR : Speaker Dependent Digit string Recognition

LPC SUBSET PERFORMANCE FOR SIMULTANEOUS SPEAKER & WORD RECOGNITION USING 3 DIGIT STRING UTTERANCES

Figure 4.13

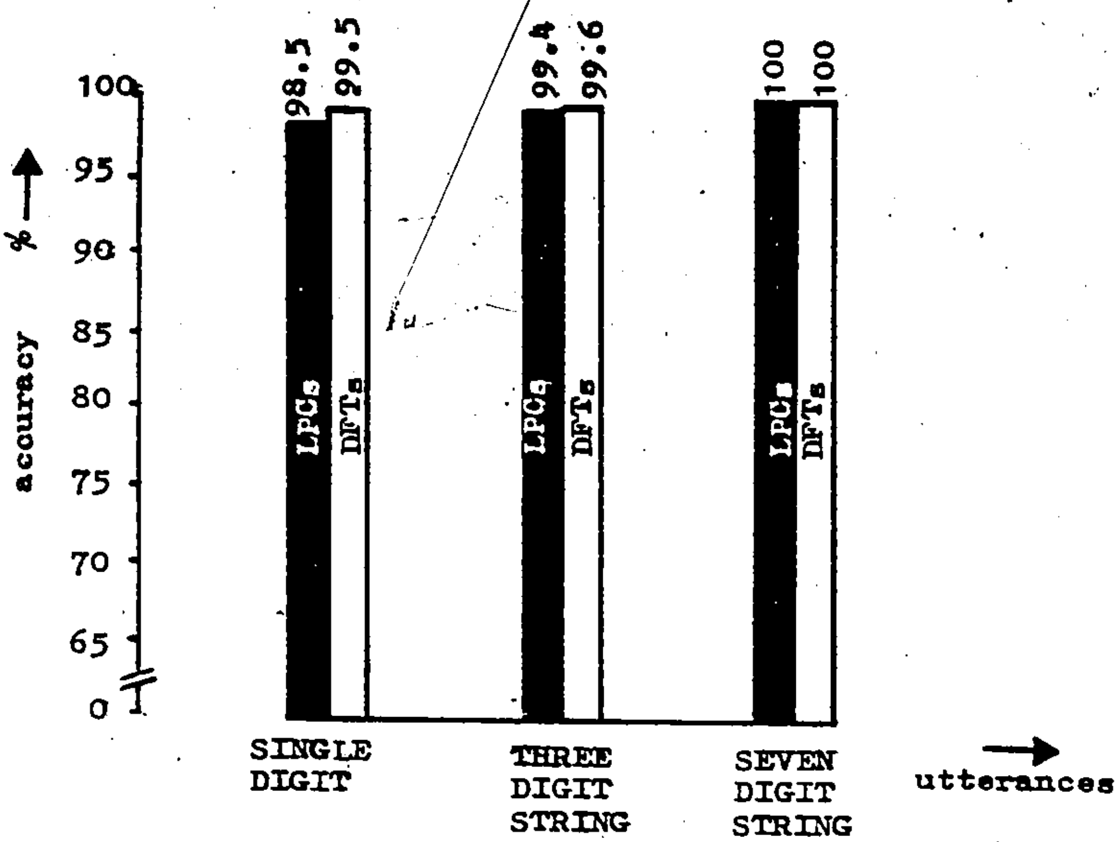




TISV : Text Independent Speaker Verification
SDDR : Speaker Dependent Digit string Recognition

LPC SUBSET PERFORMANCE FOR SIMULTANEOUS SPEAKER & WORD RECOGNITION USING 7 DIGIT STRING UTTERANCES

Figure 4.14



IMPROVEMENT IN TEXT-INDEPENDENT SPEAKER VERIFICATION WITH DIGIT STRING UTTERANCES

(FEATURE-WISE PERSPECTIVE)

Figure 4.15

Chapter V

SUMMARY AND CONCLUSIONS

In this work, the development of a combined speaker verification and digit recognition system that would operate with LPCs, orthogonal parameters (LPCs,IFSCs,DFTs) for single digit utterances were investigated. The development of a simultaneous speaker verification and digit string recognition with orthogonal parameters (LPCs,DFTs) for digit string (3-digit and 7-digit) utterances was also investigated. Effectiveness of Itakura's distance with LPCs, distance of averages and average distance with orthogonal parameters was investigated in these experiments. Also orthogonal LPC parameter subsets were investigated for speaker sensitive and speech sensitive characteristics, for single digit and digit string (3-digit, 7-digit) utterances.

The conclusions based on this work described above can be briefly stated as follows :

1. PARAMETERS FOR SIMULTANEOUS SPEAKER AND DIGIT RECOGNITION:

Orthogonal DFT parameters derived from spoken digits gave the best combined speaker and digit recognition performance. Improvement in Text Independent Speaker

Verification (TIDSV) and Speaker Dependent Digit Recognition (SDDR) obtained with orthogonal DFTs as compared to orthogonal LPCs, IFSCs is given below.

	1-digit		3-digit		7-digit	
	TIDSV	SDDR	TIDSV	SDDR	TIDSV	SDDR
LPCs	0.48	3.51	0.28	5.54	1.32	5.00
IFSCs	1.94	4.34				

Table 4.5

2. PARAMETER SETS FOR SIMULTANEOUS SPEAKER AND DIGIT RECOGNITION :

Examination of different subsets of orthogonal LPCs revealed that neither speaker verification nor digit recognition accuracies improved by using subsets of orthogonal LPCs. On the other hand the speaker verification and digit recognition accuracies were higher with the whole set 1 to 12, than those obtained with subsets. The following table summarizes the improvement of Text Independent Speaker Verification (TIDSV) and Speaker Dependent Digit Recognition (SDDR) accuracies obtained with set 1 to 12 as compared to those obtained with subsets 1 to 8, 4 to 9, 9 to 12.

	1-digit		3-digit		7-digit	
	TIDSV	SDDR	TIDSV	SDDR	TIDSV	SDDR
1-8	0.95	10.2	0.48	-0.98	-1.98	-6.06
4-9	0.85	6.1	1.26	2.50	-0.66	-0.72
9-12	0.79	4.6	1.59	8.48	6.04	6.08

Table 4.6

Neqetive entries in the case of subset 1-8 for 3 digit, 7 digit string utterances, indicate that subset 1-8 gave better performance than set 1 to 12.

3. SIMILARITY MEASURES FOR SIMULTANEOUS SPEAKER AND DIGIT RECOGNITION :

Experimental results revealed that 'average distance' as similarity measure offered the best speaker verification and digit recognition accuracies. The following table summarizes the improvement in Text Independent Speaker Verification (TIDSV) and Speaker Dependent Digit Recognition (SDDR) accuracies obtained by 'average distance' as compared to 'distance of averages'.

1-digit		3-digit		7-digit	
TIDSV	SDDR	TIDSV	SDDR	TIDSV	SDDR
0.626	4.85	0.574	0.573	2.466	29.28

Table 4.7

4. EFFECT OF ADDITIONAL PROCESSING STEPS :

Addition of pre-emphasis in the pre-processing functions, remarkably improved text dependent and text independent speaker verification accuracies. Average increase (in %) of text independent and text dependent speaker verification accuracies were 6.675% and 7.92% respectively, with pre-emphasis function.

5. EFFECT OF DIGIT STRINGS ON SPEAKER VERIFICATION ACCURACY:

Machine's ability to verify the claim of a speaker is enhanced by using digit string utterances. The following table summarizes the Text Independent Speaker Verification (TIDSV) accuracies obtained with single digit, 3 digit, 7 digit utterances.

	1-digit TIDSV	3-digit TIDSV	7-digit TIDSV
LPCs (1-8)	98.5%	99.4%	100%
DFTs	99.5%	99.6%	100%

Table 4.8

6. METHOD OF COMBINED SPEAKER AND DIGIT RECOGNITION :

After exhaustive tests and studies with different feature sets, subsets, similarity measures and combinations of these, it is found that the following approach is feasible.

1. Verify the speaker based on 7 digit sequence utterance in text independent mode, with orthogonal DFT parameters and average distance as similarity measure. Store the individual digits until the speaker's identity is confirmed.
2. Perform speaker dependent digit recognition, with orthogonal DFTs, to recognize each of the spoken digit in the sequence.

SUMMARY OF CONTRIBUTIONS

1. An algorithm for simultaneous speaker verification and digit recognition using single set of features was developed and tested.
2. Spectral parameters obtained by direct Fourier transform calculations on speech, were shown to be efficient features for simultaneous speaker and digit recognition.
3. Average distance computed across an utterance was shown to be a potential similarity measure for combined speaker and digit recognition.
4. Speaker verification accuracies more than 99% were shown to be feasible with the proposed algorithm, on digit string utterances.

FUTURE DEVELOPMENTS

In this section future directions to realize a highly efficient combined speaker and word recognition systems are discussed. Use of spectraly meaningful feature sets, new decision strategies are also stressed.

1. Use of cepstrum analysis :

Cepstrum analysis in speech research seeks to achieve deconvolution of three signals viz: impulse train, glottal impulse response, vocal tract impulse response, for voiced sounds. This deconvolution property can be utilized to advantage, to isolate vocal tract impulse response thereby localizing speaker sensitive [27] and message sensitive information in input speech. Hence cepstrum analysis could be a promising tool for a simultaneous speaker and word recognition.

2. Universal combined speaker & word recognition system :

Need for universal recognition system would be fulfilled if some technique of combining different speakers' training utterances to form universal (speaker-independent) reference tokens corresponding to each vocabulary entry.

Forming covariance matrix with different speakers' training utterances or by using well known clustering techniques [29,30,20], could be investigated for generating universal references.

3. Language-independent combined recognition system :

Language-independent recognition could be feasible by Itakura's time warping technique or with orthogonal parameters (LPCs, DFTs) with average distance as similarity measure.Here only requirement would be to to train the system with the relevant language utterances.

4. Automatic recognition of emotion, intonation, stress, health of the talker :

Thorough investigations into intra-speaker feature variations from context to context would definitely throw light on this problem. Intra-speaker speech differences are mainly caused by emotional, stress and health factors and if an automatic system is given the intelligence to detect these factors, it would be really fascinating.

5. Enhancement of decision strategy :

Addition of more references per vocabulary entry, for a designated speaker, and using K-Nearest-Neighbour rule for decision strategy would not only make the automatic recognition system context-free but also highly accurate. Intelligent pooling of results from different feature sets and subsets and parallel processing approaches would definitely enhance the speed and performance of these systems.

6. Recognition independent of environmental conditions :

Human word recognition abilities are remarkably tolerant of background noise, conversation can be understood even at a noisy party. No existing automatic recognition can approach this level of performance. Some parameter normalization techniques could be investigated to accomplish this, viz : amplitude normalization, normalization of short-time spectra with respect to long-time spectra, spectral channel contour smoothing.

7. Adoption of these recognition techniques for cursive and signature verification :

Instead of acoustic microphone, if a transducer to transform script and signatures to equivalent electrical signals is available, probably these recognition techniques could be investigated for this application.

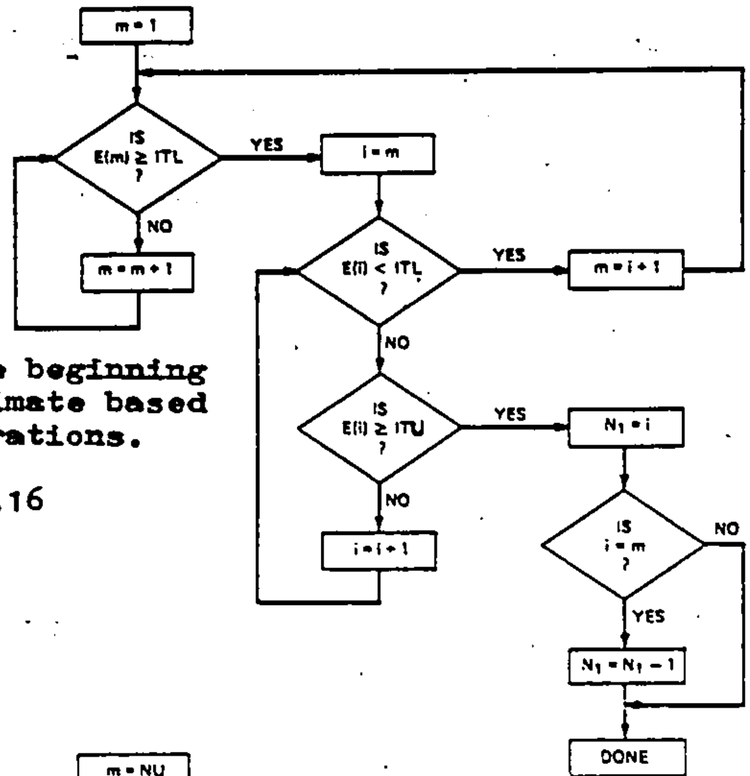
8. Real-time combined speaker and word recognition systems :

Availability of special purpose data processors capable of dealing with significantly higher data rates than general purpose computers, it would be feasible to realize a real-time AMSR system. Dynamic time alignment schemes requiring extra computation can be benefited by special purpose data processors, and availability of FFT processors, pre-processing hardware blocks, implementation of some parallel processing schemes for different feature examinations, pooling of recognition results from different parameter sets and subsets, coupled with software module for decision making, would lead to a real-time combined recognition system.

Appendix A

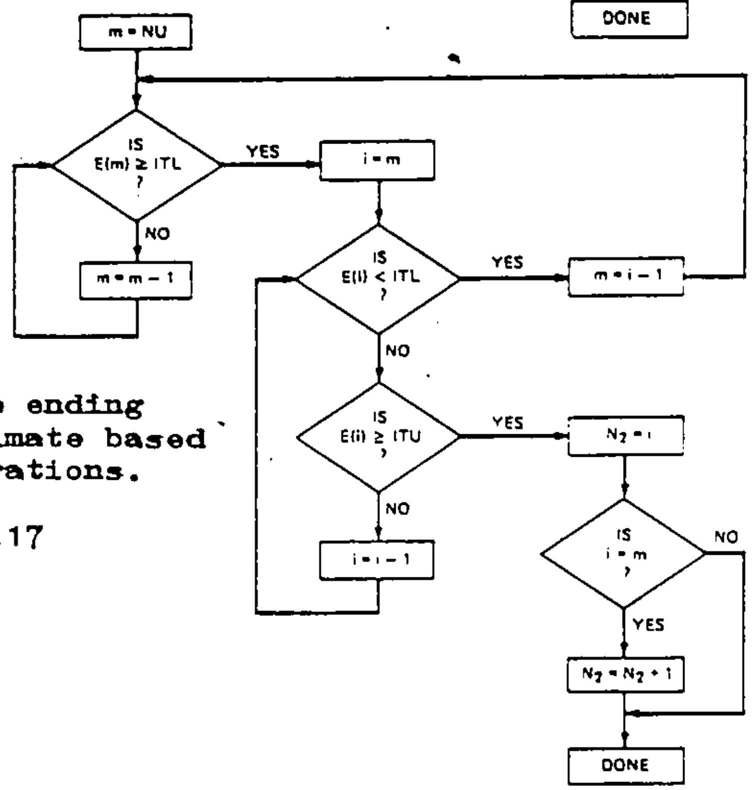
WORD END-POINT DETECTION ALGORITHM

The algorithm for finding the beginning point initial estimate is shown in figure 4.16. The algorithm begins by searching from the beginning of the interval until the lower threshold is exceeded. This point is preliminarily labeled the beginning of the utterance unless the energy falls below ITL before it rises above ITU. Should this occur, a new beginning point is obtained by finding the first point at which energy exceeds ITL, and then exceeds ITU before falling below ITL.



Flow chart for the beginning point initial estimate based on energy considerations.

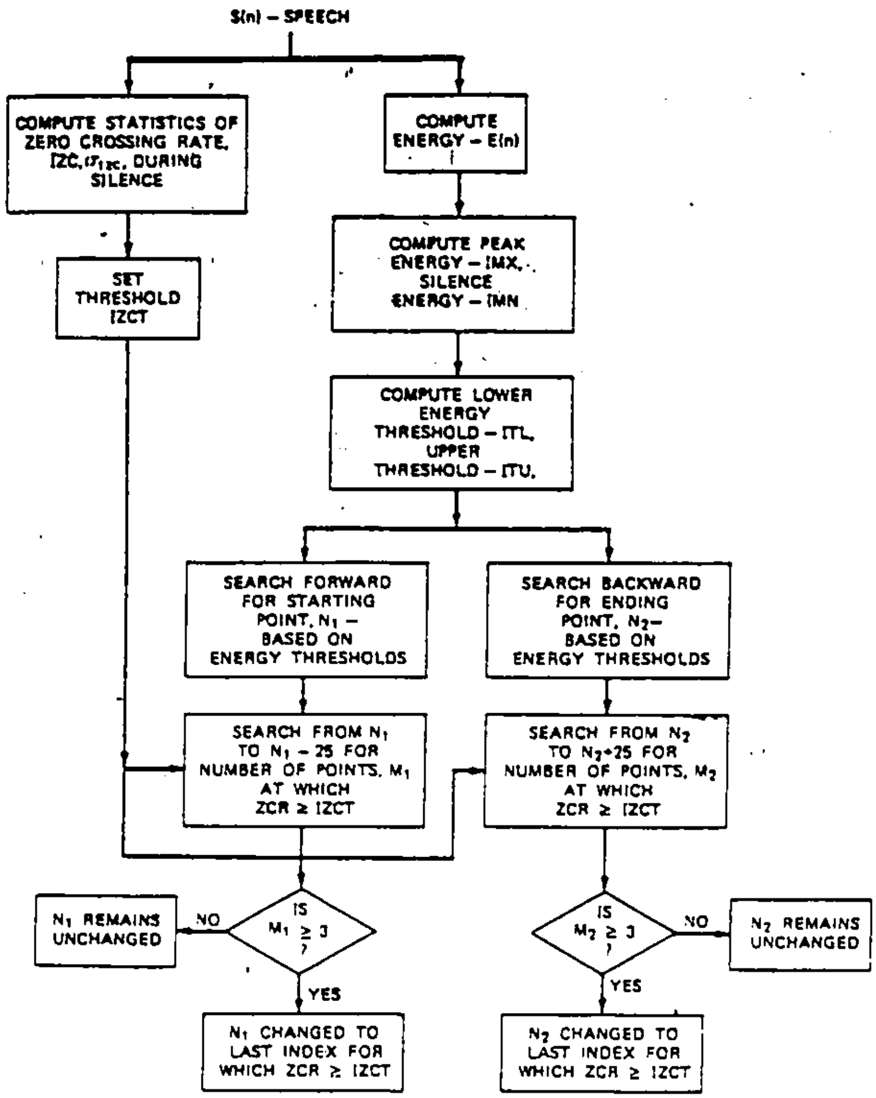
Figure 4.16



Flow chart for the ending point initial estimate based on energy considerations.

Figure 4.17

A similar algorithm as shown in figure 4.17 was used to define a preliminary estimate of the end point.The utterance within these initial estimates of begin and end points are denoted as N1 and N2 respectively. The algorithm proceeds to examine a 250 msecs interval preceding the initial beginning point, and counts the number of times the zero crossing rate exceeds the threshold IZCT. If the number of times the threshold was exceeded was three or more, the starting point was set back to the first point at which the threshold was exceeded. Otherwise the beginning point was kept at N1.A similar search procedure was used on the end point of the utterance to determine if there is unvoiced energy in the interval from N2 to 250 msecs succeeding the initial estimate of end point. The end point was readjusted based on the zero crossing test results in this interval. Figure 4.18 shows the overall flow chart for the word end-point detection algorithm, describing individual operations at each level.



<u>FLOW CHART FOR THE ENDPOINT ALGORITHM</u>

Figure 4.18

Appendix B

LPC CALCULATIONS

Prediction error :

$e(n) = s(n) - s'(n)$

where $s'(n)$ is n th predicted speech sample

$s(n)$ is n th actual speech sample

and $s'(n) = - \sum_{i=1}^{p} a_i s(n-i)$

a_i, $i=1,2,\ldots,P$ define predictor coefficients.

Mean squared prediction error :

$$E = \sum_{n=n0}^{n1} e^2(n)$$

$$= \sum_{n=n0}^{n1} [s(n) - \sum_{i=1}^{p} a_i s(n-i)]^2$$

$$= \sum_{n=n0}^{n1} [\sum_{i=0}^{p} a_i s(n-i)]^2$$

$$= \sum_{n=n0}^{n1} \sum_{i=0}^{p} \sum_{j=0}^{p} a_i s(n-i) s(n-j) a_j$$

Defining $c_{ij} = \sum_{n=n0}^{n1} s(n-i) s(n-j)$,

$$E = \sum_{i=0}^{P} \sum_{j=0}^{P} a_i c_{ij} a_j$$

$$(\partial E / \partial a_k) = 0 = 2 \sum_{i=0}^{p} a_i c_{ik}$$

P unknown predictor coefficients $[a_i]$ are obtained by solving this set of P linear simultaneous equations. c_{ik}, $i=0,1,\ldots,P$; $k=0,1,\ldots,P$ are defined from the speech data. Two methods for LPC analysis emerge out of consideration of different limits of summation and the definition of the waveform segment s(n). Covariance method is defined by setting n0=P and n1=N-1 so that the error is minimized only over the interval [P,N-1] and all N speech samples are used in calculating the covariance matrix elements c_{ik}. The autocorrelation method is defined by setting $n0=-\infty$ and $n1=\infty$ and defining s(n)=0 for n<0 and $n \geq N$. Detailed mathematical formulations are as given in the next page.

$$
\begin{aligned}
c_{ij} &= \sum_{n=-\infty}^{\infty} s(n-i)\, s(n-j) \\
&= \sum_{n=-\infty}^{\infty} s(n)\, s(n+|i-j|) \\
&= \sum_{n=0}^{N-1-|i-j|} s(n)\, s(n+|i-j|) \\
&= r(|i-j|)
\end{aligned}
$$

Covariance method :

Solve

$$\sum_{i=1}^{P} a_i c_{ij} = -c_{0j}$$

for $j = 1, 2, \ldots, P$

where

$$c_{ij} = \sum_{n=P}^{N-1} s(n-i)\, s(n-j)$$

with error

$$e(n) = \sum_{i=0}^{P} a_i s(n-i) \qquad (a_0 = 1)$$

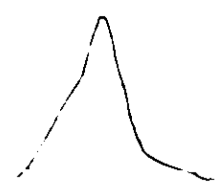
$n = P, P+1, \ldots, N-1$

Autocorrelation method :

Solve

$$\sum_{i=1}^{P} a_i r(|i-j|) = -r(j)$$

for $j = 1, 2, \ldots, P$



where

$$r(l) = \sum_{n=0}^{N-1-l} s(n)\, s(n+l) \qquad (l \geq 0)$$

with error

$$e(n) = \sum_{i=0}^{P} a_i s(n-i) \qquad (a_0 = 1)$$

for $n = 0, 1, \ldots, N+P-1$

Appendix C

DYNAMIC TIME WARPING

Speech can be represented as a sequence of feature vectors :

$$C = a1, a2, \ldots, a_m, \ldots a_M$$

$$D = b1, b2, \ldots, b_n, \ldots b_N$$

Consider the problem of eliminating timing differences between two speech patterns. In order to clarify the nature of time-axis fluctuations or timing differences, let us consider a 'm - n' plane, where patterns C and D are developed along m th axis and n th axis respectively.

The timing differences between two utterances can be depicted by a sequence of mapping points

$$F = w(1), w(2), \ldots, w(k), \ldots w(K)$$

$$\text{where } w(k) = (m(k), n(k))$$

This sequence can be considered to represent a function which approximately realizes a mapping from the time-axis of a pattern C onto that of a pattern D. Locus of these sequence of points is also called as 'warping function'.

When there is no timing difference between these two patterns, the warping function coincides with the diagonal line $m = n$. It deviates further from the diagonal line as the timing difference grows as shown in the figure 4.19. The warping function becomes a straight line of slope 2 and 1/2, when the timing of the test utterance becomes half and twice that of reference utterance, as shown in the figure 4.20.

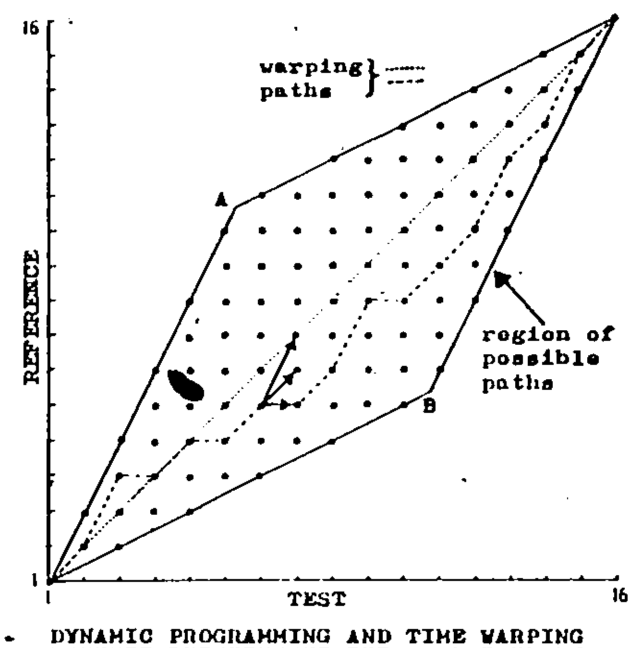


Figure 4.19

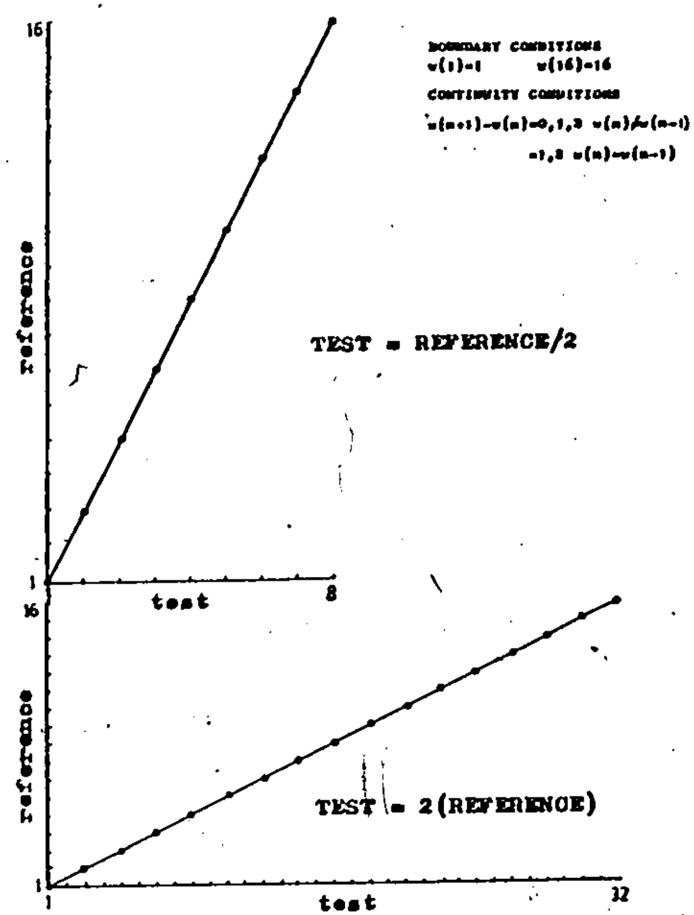


Figure 4.20

The warping function $w(k)$ must satisfy a set of boundary conditions at the endpoints of the utterance. Typical assumption is that both the initial and final points of the test and reference utterances are in time alignment i.e.

$$M1=w(N1) \text{----------} (1)\ (a)$$

$$M2=w(N2) \text{----------} (1)\ (b)$$

A more sophisticated approach to time-warping is to constrain the warping function to satisfy a set of continuity conditions :

$$w(n+1)-w(n)=0,1,2 \quad \text{for } w(n)\neq w(n-1) \text{--------} (2)\ (a)$$

$$=1,2 \quad \text{for } w(n)=w(n-1) \text{--------} (2)\ (b)$$

These equations require that $w(n)$ be monotonicaly increasing with a maximum slope of 2 and minimum slope of 1/2 except when the slope at the preceding frame was zero, in which case the minimum slope was made equal to 1. The boundary conditions of (1) and continuity conditions of (2) constrain the warping function to lie within a parallelogram in the (n,m) plane as shown in the figure 4.20. The vertices of the parallelogram A and B are calculated as the intersections of the lines

$m-1 = 2(n-1)$ and $m-M = (n-N)/2$ for point A

$m-1 = (1/2)(n-1)$ and $m-M = 2(n-N)$ for point B.

A similarity measure (Itakura's distance) must be defined for every pair of points (n,m) within the parallelogram. An optimum path 'w' can be calculated by means of dynamic programming tool. Multi-level minimization at different stretchings and shrinkings of time-axis of reference patterns, and simultaneous accumulation of distances along the warping path is achieved recursively as given in this formula.

$$D(n+1,m) = d(n+1,m) + \min(D(n,m) \cdot q(n,m),\ D(n,m-1),\ D(n,m-2))$$

where $q(n,m) = 1$ for $w(n) \neq w(n-1)$

$= \infty$ for $w(n) = w(n-1)$

and $d(n,m)$ and $D(n,m)$ are the local and accumulated distance between n th frame of the test and m th frame of the reference.

The global distance between the test and reference utterance is the final solution $D = D(N,M)$ at final point in the recursion.

Appendix D

TABULATION OF RESULTS (SINGLE DIGIT EXPERIMENTS)

TEXT DEPENDENT SPEAKER VERIFICATION EXPERIMENTS

TEXT-DEPENDENT SPEAKER VERIFICATION USING LINEAR PREDICTION COEFFICIENTS : 9 to 12

Table 4.9.1

'Average distance'

		SPOKEN DIGITS									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	0	13	0	0	0	3	0	1	0
	Recognition accuracy %	100	100	62.9	100	100	100	91.4	100	97.1	100
Speaker 'A'	Total errors	0	0	12	0	8	1	0	1	0	0
	Recognition accuracy %	100	100	65.7	100	77.1	97.1	100	97.1	100	100
Speaker 'J'	Total errors	1	0	0	2	0	1	3	0	0	2
	Recognition accuracy %	97.1	100	100	94.3	100	97.1	91.4	100	100	94.3
Speaker 'P'	Total errors	2	9	0	0	9	4	9	5	7	1
	Recognition accuracy %	94.3	74.3	100	100	74.3	88.6	74.3	85.7	80	97.1
Speaker 'E'	Total errors	0	0	0	0	0	1	2	8	1	0
	Recognition accuracy %	100	100	100	100	100	97.1	94.3	77.1	97.1	100
Speaker 'R'	Total errors	2	3	1	3	4	4	2	0	0	4
	Recognition accuracy %	94.3	91.4	97.1	91.4	88.6	88.6	94.3	100	100	88.6
Speaker 'W'	Total errors	0	0	1	4	3	1	1	10	1	3
	Recognition accuracy %	100	100	97.1	88.6	91.4	97.1	97.1	71.4	97.1	91.4

OVERALL TOTAL ERRORS	268
OVERALL RECOGNITION ACCURACY IN %	89

'Distance of averages'

		SPOKEN DIGITS									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	0	15	0	2	0	7	15	1	1
	Recognition accuracy %	100	100	57.1	100	94.3	100	80	57.1	97.1	97.1
Speaker 'A'	Total errors	2	8	18	0	4	4	1	4	0	0
	Recognition accuracy %	94.3	77.1	48.6	100	88.6	88.6	97.1	88.6	100	100
Speaker 'J'	Total errors	2	0	0	2	1	2	1	0	0	8
	Recognition accuracy %	94.3	100	100	94.3	97.1	94.3	97.1	100	100	77.1
Speaker 'P'	Total errors	3	10	0	0	9	0	8	6	4	12
	Recognition accuracy %	91.4	71.4	100	100	74.3	100	77.1	82.9	88.6	65.7
Speaker 'E'	Total errors	0	0	0	0	2	1	4	4	2	0
	Recognition accuracy %	100	100	100	100	94.3	97.1	88.6	88.6	94.3	100
Speaker 'R'	Total errors	0	1	0	5	1	3	1	0	2	10
	Recognition accuracy %	100	97.1	100	85.7	97.1	91.4	97.1	100	94.3	71.4
Speaker 'W'	Total errors	0	0	2	3	6	7	3	5	2	11
	Recognition accuracy %	100	100	94.3	91.4	82.9	80	91.4	85.7	94.3	68.6

OVERALL TOTAL ERRORS	225
OVERALL RECOGNITION ACCURACY IN %	90.8

TEXT-DEPENDENT SPEAKER VERIFICATION USING LINEAR PREDICTION COEFFICIENTS: 4 to 9

Table 4.9.2

'Average distance'

					SPOKEN DIGITS						
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	0	6	0	0	0	6	0	3	0
	Recognition accuracy %	100	100	82.9	100	100	100	82.9	100	91.4	100
Speaker 'A'	Total errors	0	3	13	6	9	2	3	1	2	6
	Recognition accuracy %	100	91.4	62.9	82.9	74.3	94.3	91.4	97.1	94.3	82.9
Speaker 'J'	Total errors	0	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	100
Speaker 'P'	Total errors	3	3	3	0	3	5	2	6	2	5
	Recognition accuracy %	91.4	91.4	91.4	100	91.4	85.7	94.3	82.9	94.3	85.7
Speaker 'E'	Total errors	0	1	0	0	6	0	2	0	0	0
	Recognition accuracy %	100	97.1	100	100	82.9	100	94.3	100	100	100
Speaker 'R'	Total errors	2	0	0	0	1	1	1	8	1	5
	Recognition accuracy %	94.3	100	100	100	97.1	97.1	97.1	77.1	97.1	85.7
Speaker 'V'	Total errors	1	2	2	3	1	1	5	16	2	2
	Recognition accuracy %	97.1	94.3	94.3	91.4	97.1	97.1	85.7	54.3	94.3	94.3

OVERALL TOTAL ERRORS	155
OVERALL RECOGNITION ACCURACY IN %	93.7

'Distance of averages'

					SPOKEN DIGITS						
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	0	13	0	0	1	6	0	2	3
	Recognition accuracy %	100	100	62.9	100	100	97.1	82.9	100	94.3	91.4
Speaker 'A'	Total errors	2	3	15	7	9	5	2	0	7	1
	Recognition accuracy %	94.3	91.4	57.1	80	74.3	85.7	94.3	100	80	97.1
Speaker 'J'	Total errors	0	0	0	0	1	2	0	3	1	0
	Recognition accuracy %	100	100	100	100	97.1	94.3	100	91.4	97.1	100
Speaker 'P'	Total errors	1	4	2	0	2	2	2	4	1	2
	Recognition accuracy %	97.1	88.6	94.3	100	94.3	94.3	94.3	88.6	97.1	94.3
Speaker 'E'	Total errors	0	4	0	0	7	2	2	3	0	5
	Recognition accuracy %	100	88.6	100	100	80	94.3	94.3	91.4	100	85.7
Speaker 'R'	Total errors	0	0	0	0	0	0	2	1	0	3
	Recognition accuracy %	100	100	100	100	100	100	94.3	97.1	100	91.4
Speaker 'V'	Total errors	4	5	1	2	7	4	6	4	2	4
	Recognition accuracy %	88.6	85.7	97.1	94.3	80	88.6	82.9	88.6	94.3	88.6

OVERALL TOTAL ERRORS	172
OVERALL RECOGNITION ACCURACY IN %	93

TEXT-DEPENDENT SPEAKER VERIFICATION USING LINEAR PREDICTION COEFFICIENTS : 1 to 8

Table 4.9.3

'Average distance'

Speaker		Spoken digits 0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	0	4	0	0	2	9	3	3	0
	Recognition accuracy %	100	100	88.6	100	100	94.3	74.3	91.4	91.4	100
Speaker 'A'	Total errors	0	5	7	6	15	5	5	0	2	3
	Recognition accuracy %	100	85.7	80	82.9	57.1	85.7	85.7	100	94.3	91.4
Speaker 'J'	Total errors	0	0	0	0	0	0	0	2	1	0
	Recognition accuracy %	100	100	100	100	100	100	100	94.3	97.1	100
Speaker 'P'	Total errors	1	2	1	0	5	10	6	4	9	3
	Recognition accuracy %	97.1	94.3	97.1	100	85.7	71.4	82.9	88.6	74.3	91.4
Speaker 'E'	Total errors	3	0	0	0	0	1	1	0	1	4
	Recognition accuracy %	91.4	100	100	100	100	97.1	97.1	100	97.1	88.6
Speaker 'R'	Total errors	3	0	0	0	0	1	1	0	1	4
	Recognition accuracy %	91.4	100	100	100	100	97.1	97.1	100	97.1	88.6
Speaker 'W'	Total errors	0	3	2	2	1	1	3	15	0	2
	Recognition accuracy %	100	91.4	94.3	94.3	97.1	97.1	91.4	57.1	100	94.3

OVERALL TOTAL ERRORS	158
OVERALL RECOGNITION ACCURACY IN %	93.6

'Distance of averages'

Speaker		Spoken digits 0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	0	9	0	0	4	17	4	2	0
	Recognition accuracy %	100	100	74.3	100	100	88.6	51.4	88.6	94.3	100
Speaker 'A'	Total errors	0	5	9	8	16	9	2	0	6	1
	Recognition accuracy %	100	85.7	74.3	77.1	54.3	74.3	94.3	100	82.9	97.1
Speaker 'J'	Total errors	0	2	0	2	1	3	0	3	0	0
	Recognition accuracy %	100	94.3	100	94.3	97.1	91.4	100	91.4	100	100
Speaker 'P'	Total errors	2	2	0	0	4	7	2	0	1	3
	Recognition accuracy %	94.3	94.3	100	100	88.6	80	94.3	100	97.1	91.4
Speaker 'E'	Total errors	0	3	0	0	2	2	2	0	1	2
	Recognition accuracy %	100	91.4	100	100	94.3	94.3	94.3	100	97.1	94.3
Speaker 'R'	Total errors	2	0	0	0	0	4	1	0	0	2
	Recognition accuracy %	94.3	100	100	100	100	88.6	97.1	100	100	94.3
Speaker 'W'	Total errors	2	5	3	5	9	7	3	9	2	6
	Recognition accuracy %	94.3	85.7	91.4	85.7	74.3	80	91.4	74.3	94.3	82.9

OVERALL TOTAL ERRORS	196
OVERALL RECOGNITION ACCURACY IN %	92

TEXT-DEPENDENT SPEAKER VERIFICATION USING LINEAR PREDICTION COEFFICIENTS : 1 to 12

Table 4.9.4

'Average distance'

		SPOKEN DIGITS									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	0	9	0	0	0	5	0	2	0
	Recognition accuracy %	100	100	74.3	100	100	100	85.7	100	94.3	100
Speaker 'A'	Total errors	0	2	9	0	11	1	0	0	1	0
	Recognition accuracy %	100	94.3	74.3	100	68.6	97.1	100	100	97.1	100
Speaker 'J'	Total errors	0	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	100
Speaker 'P'	Total errors	0	4	0	0	1	7	5	2	6	1
	Recognition accuracy %	100	88.6	100	100	97.1	80	85.7	94.3	82.9	97.1
Speaker 'E'	Total errors	0	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	100
Speaker 'R'	Total errors	2	2	0	2	2	0	0	0	0	3
	Recognition accuracy %	94.3	94.3	100	94.3	94.3	100	100	100	100	91.4
Speaker 'W'	Total errors	0	1	0	4	1	0	1	12	1	2
	Recognition accuracy %	100	97.1	100	88.6	97.1	100	97.1	65.7	97.1	94.3

OVERALL TOTAL ERRORS	100
OVERALL RECOGNITION ACCURACY IN %	95.92

'Distance of averages'

		SPOKEN DIGITS									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	0	4	0	0	0	6	1	1	0
	Recognition accuracy %	100	100	88.6	100	100	100	82.9	97.1	97.1	100
Speaker 'A'	Total errors	0	4	12	2	12	4	1	0	3	0
	Recognition accuracy %	100	88.6	65.7	94.3	65.7	88.6	97.1	100	91.4	100
Speaker 'J'	Total errors	0	0	0	0	0	3	0	0	0	4
	Recognition accuracy %	100	100	100	100	100	91.4	100	100	100	88.6
Speaker 'P'	Total errors	1	1	0	0	2	3	1	0	1	9
	Recognition accuracy %	97.1	97.1	100	100	94.3	91.4	97.1	100	97.1	74.3
Speaker 'E'	Total errors	0	0	0	0	2	2	0	1	1	0
	Recognition accuracy %	100	100	100	100	94.3	94.3	100	97.1	97.1	100
Speaker 'R'	Total errors	0	0	0	2	0	0	0	0	0	3
	Recognition accuracy %	100	100	100	94.3	100	100	100	100	100	91.4
Speaker 'W'	Total errors	0	2	0	0	1	3	4	3	2	6
	Recognition accuracy %	100	94.3	100	100	97.1	91.4	88.6	91.4	94.3	82.9

OVERALL TOTAL ERRORS	107
OVERALL RECOGNITION ACCURACY IN %	95.6

TEXT-DEPENDENT SPEAKER VERIFICATION USING INVERSE FILTER SPECTRAL COEFFICIENTS 1 to 32

Table 4.9.5

'Average distance'

		SPOKEN DIGITS									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	5	4	2	9	2	32	2	18	0
	Recognition accuracy %	100	85.7	88.6	94.3	74.3	94.3	8.6	94.3	48.6	100
Speaker 'A'	Total errors	0	2	1	0	4	4	0	0	0	0
	Recognition accuracy %	100	94.3	97.1	100	88.6	88.6	100	100	100	100
Speaker 'J'	Total errors	0	0	0	0	0	1	0	4	0	0
	Recognition accuracy %	100	100	100	100	100	97.1	100	88.6	100	100
Speaker 'P'	Total errors	1	4	0	0	1	4	1	1	4	19
	Recognition accuracy %	97.1	88.6	100	100	97.1	88.6	97.1	97.1	88.6	45.7
Speaker 'E'	Total errors	0	0	0	2	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	94.3	100	100	100	100	100	100
Speaker 'R'	Total errors	0	1	0	1	2	1	0	2	0	6
	Recognition accuracy %	100	97.1	100	97.1	94.3	97.1	100	94.3	100	82.9
Speaker 'V'	Total errors	0	0	1	1	1	0	2	13	0	1
	Recognition accuracy %	100	100	97.1	97.1	97.1	100	94.3	62.9	100	97.1

OVERALL TOTAL ERRORS	159
OVERALL RECOGNITION ACCURACY IN %	93.5

'Distance of averages'

		SPOKEN DIGITS									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	7	11	0	1	0	34	3	3	0
	Recognition accuracy %	100	80	68.6	100	97.1	100	2.86	91.4	91.4	100
Speaker 'A'	Total errors	0	0	0	0	5	3	3	0	0	0
	Recognition accuracy %	100	100	100	100	85.7	91.4	91.4	100	100	100
Speaker 'J'	Total errors	2	0	0	0	1	1	1	7	0	5
	Recognition accuracy %	94.3	100	100	100	97.1	97.1	97.1	80	100	85.7
Speaker 'P'	Total errors	1	0	0	0	1	2	3	2	0	4
	Recognition accuracy %	97.1	100	100	100	97.1	94.3	91.4	94.3	100	88.6
Speaker 'E'	Total errors	0	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	100
Speaker 'R'	Total errors	0	1	0	0	2	2	0	0	0	3
	Recognition accuracy %	100	97.1	100	100	94.3	94.3	100	100	100	91.4
Speaker 'V'	Total errors	3	0	0	0	1	0	6	4	0	2
	Recognition accuracy %	91.4	100	100	100	97.1	100	82.9	88.6	100	94.3

OVERALL TOTAL ERRORS	124
OVERALL RECOGNITION ACCURACY IN %	94.9

TEXT-DEPENDENT SPEAKER VERIFICATION USING DIRECT FOURIER TRANSFORM OF SPEECH: 1 to 32

Table 4.9.6

'Average distance'

		Spoken digits									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	0	3	0	0	6	0	2	0	0
	Recognition accuracy %	100	100	91.4	100	100	82.9	100	94.3	100	100
Speaker 'A'	Total errors	0	0	4	0	4	2	3	0	5	3
	Recognition accuracy %	100	100	88.6	100	88.6	94.3	91.4	100	85.7	91.4
Speaker 'J'	Total errors	0	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	100
Speaker 'P'	Total errors	0	0	0	0	4	1	2	0	5	15
	Recognition accuracy %	100	100	100	100	88.6	97.1	94.3	100	85.7	57.1
Speaker 'E'	Total errors	0	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	100
Speaker 'R'	Total errors	0	0	0	0	1	15	1	0	0	0
	Recognition accuracy %	100	100	100	100	97.1	57.1	97.1	100	100	100
Speaker 'W'	Total errors	1	0	0	6	0	1	2	2	2	2
	Recognition accuracy %	97.1	100	100	82.9	100	97.1	94.3	94.3	94.3	94.3

OVERALL TOTAL ERRORS	82
OVERALL RECOGNITION ACCURACY IN %	96.7

'Distance of averages'

		Spoken digits									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	11	9	9	8	9	12	23	0	0
	Recognition accuracy %	100	68.6	74.3	74.3	77.1	74.3	65.7	34.3	100	100
Speaker 'A'	Total errors	0	0	4	0	10	5	2	7	4	5
	Recognition accuracy %	100	100	88.6	100	71.4	85.7	94.3	80	88.6	85.7
Speaker 'J'	Total errors	5	4	0	0	0	0	0	6	0	2
	Recognition accuracy %	85.7	88.6	100	100	100	100	100	82.9	100	94.3
Speaker 'P'	Total errors	0	0	0	2	4	0	3	0	3	8
	Recognition accuracy %	100	100	100	94.3	88.6	100	91.4	100	91.4	77.1
Speaker 'E'	Total errors	0	0	0	0	0	2	1	2	0	4
	Recognition accuracy %	100	100	100	100	100	94.3	97.1	94.3	100	88.6
Speaker 'R'	Total errors	0	0	0	0	1	20	1	0	0	0
	Recognition accuracy %	100	100	100	100	97.1	42.9	97.1	100	100	100
Speaker 'W'	Total errors	2	0	0	10	1	1	13	9	5	4
	Recognition accuracy %	94.3	100	100	71.4	97.1	97.1	62.9	74.3	85.7	88.6

OVERALL TOTAL ERRORS	231
OVERALL RECOGNITION ACCURACY IN %	90.57

Appendix E

TABULATION OF RESULTS (SINGLE DIGIT EXPERIMENTS)

COMBINED TEXT INDEPENDENT SPEAKER VERIFICATION AND

SPEAKER DEPENDENT DIGIT RECOGNITION

USING LINEAR PREDICTION COEFFICIENTS : 1 to 12 : Itakura's distance

Table 4.10.1

TEXT-INDEPENDENT SPEAKER VERIFICATION

		SPOKEN DIGITS									
		0	1	2	3	4	5	6	7	8	9
Speaker 'A'	Total errors	17	8	0	26	43	11	49	9	22	6
	Recognition accuracy %	83.3	92.2	100	74.5	57.8	89.2	52	91.2	78.4	94.1
Speaker 'J'	Total errors	32	2	23	0	5	15	20	27	53	33
	Recognition accuracy %	68.6	98.1	77.5	100	95.1	85.3	80.4	73.5	48	67.7
Speaker 'P'	Total errors	10	17	6	35	15	1	36	51	7	1
	Recognition accuracy %	90.2	83.3	94.1	65.8	85.3	99	64.7	50	93.1	99
Speaker 'E'	Total errors	3	4	3	13	0	8	5	0	9	0
	Recognition accuracy %	97.1	96.1	97.1	87.3	100	92.2	95.1	100	91.2	100
Speaker 'R'	Total errors	14	19	0	1	29	41	3	8	21	2
	Recognition accuracy %	86.3	81.4	100	99	71.6	59.8	97.1	92.2	79.4	98.1
Speaker 'W'	Total errors	0	0	0	9	0	0	10	4	0	13
	Recognition accuracy %	100	100	100	91.2	100	100	90.2	96.1	100	87.3

OVERALL TOTAL ERRORS	799
OVERALL RECOGNITION ACCURACY IN %	86.9

SPEAKER-DEPENDENT DIGIT RECOGNITION

		SPOKEN DIGITS									
		0	1	2	3	4	5	6	7	8	9
Speaker 'A'	Total errors	0	0	0	0	0	7	7	0	0	4
	Recognition accuracy %	100	100	100	100	100	65	65	100	100	80
Speaker 'J'	Total errors	0	0	0	0	0	0	0	1	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	95	100	100
Speaker 'P'	Total errors	0	4	0	0	0	0	0	2	11	0
	Recognition accuracy %	100	80	100	100	100	100	100	90	45	100
Speaker 'E'	Total errors	0	0	0	0	0	8	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	60	100	100	100	100
Speaker 'R'	Total errors	0	0	0	0	0	2	4	0	0	0
	Recognition accuracy %	100	100	100	100	100	90	80	100	100	100
Speaker 'W'	Total errors	0	0	10	4	0	0	0	0	0	0
	Recognition accuracy %	100	100	50	80	100	100	100	100	100	100

OVERALL TOTAL ERRORS	64
OVERALL RECOGNITION ACCURACY IN %	94.7

LINEAR PREDICTION COEFFICIENTS : 9 to 12 : 'Average distance'

Table 4.10.2

TEXT-INDEPENDENT SPEAKER VERIFICATION

					Spoken digits						
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	1	0	48	2	0	1	3	2	1	0
	Recognition accuracy %	99.7	100	84.3	99.3	100	99.7	99	99.3	99.7	100
Speaker 'A'	Total errors	16	4	137	2	10	1	0	3	0	5
	Recognition accuracy %	94.8	98.7	55.1	99.3	96.7	99.7	100	99	100	96.4
Speaker 'J'	Total errors	5	2	0	2	0	1	3	0	0	5
	Recognition accuracy %	96.4	99.3	100	99.3	100	99.7	99	100	100	96.4
Speaker 'P'	Total errors	12	12	2	3	9	5	9	8	9	6
	Recognition accuracy %	96.1	96.1	99.3	99	97.1	98.4	97.1	97.4	97.4	98
Speaker 'E'	Total errors	2	0	0	0	0	1	2	13	2	0
	Recognition accuracy %	99.3	100	100	100	100	99.7	99.3	95.7	99.3	100
Speaker 'R'	Total errors	20	6	2	6	4	4	2	5	0	9
	Recognition accuracy %	93.5	98	99.3	98	98.7	98.7	99.3	98.4	100	97.1
Speaker 'W'	Total errors	4	0	2	7	3	1	1	11	1	3
	Recognition accuracy %	98.7	100	99.3	97.7	99	99.7	99.7	96.4	99.7	99

OVERALL TOTAL ERRORS	440
OVERALL RECOGNITION ACCURACY IN %	97.9

SPEAKER-DEPENDENT DIGIT RECOGNITION

					Spoken digits						
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	4	38	12	0	17	7	1	2	1
	Recognition accuracy %	100	92	24	76	100	66	86	98	96	98
Speaker 'A'	Total errors	29	0	45	9	5	5	4	16	0	4
	Recognition accuracy %	42	100	10	82	90	90	92	68	100	92
Speaker 'J'	Total errors	2	0	0	5	0	4	0	0	0	3
	Recognition accuracy %	96	100	100	90	100	92	100	100	100	94
Speaker 'P'	Total errors	0	1	1	4	0	0	0	4	1	4
	Recognition accuracy %	100	98	98	92	100	100	100	92	98	92
Speaker 'E'	Total errors	0	0	0	0	0	1	0	7	0	0
	Recognition accuracy %	100	100	100	100	100	98	100	86	100	100
Speaker 'R'	Total errors	6	2	0	11	0	0	2	2	0	0
	Recognition accuracy %	88	96	100	78	100	100	96	96	100	100
Speaker 'W'	Total errors	4	0	1	24	14	0	0	30	0	4
	Recognition accuracy %	92	100	98	52	72	100	100	40	100	92

OVERALL TOTAL ERRORS	336
OVERALL RECOGNITION ACCURACY IN %	90.4

USING LINEAR PREDICTION COEFFICIENTS : 4 to 9 : 'Average distance'

Table 4.10.3

TEXT-INDEPENDENT SPEAKER VERIFICATION

Speaker		0	1	2	SPOKEN DIGITS 3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	0	20	2	0	0	8	0	3	0
	Recognition accuracy %	100	100	93.4	99.3	100	100	97.4	100	99	100
Speaker 'A'	Total errors	2	3	135	21	15	2	3	1	3	12
	Recognition accuracy %	99.3	99	55.7	93.1	95.1	99.3	99	99.7	99	96.1
Speaker 'J'	Total errors	2	2	0	2	1	0	0	0	0	0
	Recognition accuracy %	99.3	99.3	100	99.3	99.7	100	100	100	100	100
Speaker 'P'	Total errors	10	2	3	8	6	6	3	12	2	5
	Recognition accuracy %	96.7	99.3	99	97.4	98	98	99	96.1	99.3	98.4
Speaker 'E'	Total errors	2	4	0	0	6	4	2	2	0	0
	Recognition accuracy %	99.3	98.7	100	100	98	98.7	99.3	99.3	100	100
Speaker 'R'	Total errors	26	0	0	0	1	6	2	16	1	5
	Recognition accuracy %	91.5	100	100	100	99.7	98	99.3	94.8	99.7	98.4
Speaker 'Y'	Total errors	2	4	2	2	1	2	2	10	2	4
	Recognition accuracy %	99.3	98.7	99.3	99.3	99.7	99.3	99.3	96.7	99.3	98.7

OVERALL TOTAL ERRORS	401
OVERALL RECOGNITION ACCURACY IN %	98.12

SPEAKER-DEPENDENT DIGIT RECOGNITION

Speaker		0	1	2	SPOKEN DIGITS 3	4	5	6	7	8	9
Speaker 'S'	Total errors	7	10	25	9	1	4	9	0	8	6
	Recognition accuracy %	86	80	50	82	98	92	82	100	84	88
Speaker 'A'	Total errors	1	5	44	2	13	2	3	14	1	6
	Recognition accuracy %	98	90	12	96	74	96	94	72	98	88
Speaker 'J'	Total errors	0	0	0	3	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	94	100	100	100	100	100	100
Speaker 'P'	Total errors	12	0	0	6	3	10	2	12	0	0
	Recognition accuracy %	76	100	100	88	94	80	96	76	100	100
Speaker 'E'	Total errors	3	7	0	4	22	0	2	3	0	0
	Recognition accuracy %	94	86	100	92	56	100	96	94	100	100
Speaker 'R'	Total errors	5	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	90	100	100	100	100	100	100	100	100	100
Speaker 'Y'	Total errors	11	4	6	5	15	7	12	20	0	1
	Recognition accuracy %	78	92	88	90	70	86	76	60	100	98

OVERALL TOTAL ERRORS	345
OVERALL RECOGNITION ACCURACY IN %	90.1

USING LINEAR PREDICTION COEFFICIENTS : 1 to 8 : 'Average distance'

Table 4.10.4

TEXT-INDEPENDENT SPEAKER VERIFICATION

		Spoken digits 0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	1	7	1	0	1	11	3	3	0
	Recognition accuracy %	100	99.7	97.7	99.7	100	99.7	96.4	99	99	100
Speaker 'A'	Total errors	2	8	39	4	59	10	5	0	2	4
	Recognition accuracy %	99.3	97.4	87.2	98.7	80.7	96.7	98.4	100	99.3	98.7
Speaker 'J'	Total errors	0	2	0	0	0	0	0	2	2	0
	Recognition accuracy %	100	99.3	100	100	100	100	100	99.3	99.3	100
Speaker 'P'	Total errors	8	3	8	5	7	21	5	12	9	3
	Recognition accuracy %	97.4	99	97.4	98.4	97.7	93.1	98.4	96.1	97.1	99
Speaker 'E'	Total errors	2	2	0	0	2	1	2	0	0	0
	Recognition accuracy %	99.3	99.3	100	100	99.3	99.7	99.3	100	100	100
Speaker 'R'	Total errors	5	0	0	0	0	3	1	2	1	4
	Recognition accuracy %	98.4	100	100	100	100	99	99.7	99.3	99.7	98.7
Speaker 'W'	Total errors	1	2	4	3	1	2	20	21	0	3
	Recognition accuracy %	99.7	99.3	98.7	99	99.7	99.3	93.4	93.1	100	99

OVERALL TOTAL ERRORS	326
OVERALL RECOGNITION ACCURACY IN %	98.5

SPEAKER-DEPENDENT DIGIT RECOGNITION

		Spoken digits 0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	5	16	14	4	4	8	6	2	10	2
	Recognition accuracy %	90	68	72	92	92	84	88	96	80	96
Speaker 'A'	Total errors	0	7	24	1	16	12	0	0	1	4
	Recognition accuracy %	100	86	52	98	68	76	100	100	98	92
Speaker 'J'	Total errors	0	3	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	94	100	100	100	100	100	100	100	100
Speaker 'P'	Total errors	21	0	1	0	11	9	0	7	0	0
	Recognition accuracy %	58	100	98	100	78	82	100	86	100	100
Speaker 'E'	Total errors	5	8	0	3	9	0	1	3	0	0
	Recognition accuracy %	90	84	100	94	82	100	98	94	100	100
Speaker 'R'	Total errors	1	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	98	100	100	100	100	100	100	100	100	100
Speaker 'W'	Total errors	4	9	4	1	13	5	11	12	0	1
	Recognition accuracy %	92	82	92	98	74	90	78	76	100	98

OVERALL TOTAL ERRORS	278
OVERALL RECOGNITION ACCURACY IN %	92.1

USING LINEAR PREDICTION COEFFICIENTS : 1 to 12 : 'Average distance'

Table 4.10.5

TEXT-INDEPENDENT SPEAKER VERIFICATION

		SPOKEN DIGITS									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	0	17	1	0	0	5	0	2	0
	Recognition accuracy %	100	100	94.4	99.6	100	100	98.3	100	99.3	100
Speaker 'A'	Total errors	5	2	75	7	13	1	0	1	1	2
	Recognition accuracy %	98.3	99.3	75.4	97.7	95.7	99.6	100	99.6	99.6	99.3
Speaker 'J'	Total errors	1	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	99.6	100	100	100	100	100	100	100	100	100
Speaker 'P'	Total errors	5	7	7	1	2	4	5	5	6	1
	Recognition accuracy %	98.3	97.7	97.7	99.6	99.3	98.7	98.3	98.3	98	99.6
Speaker 'E'	Total errors	0	0	0	0	0	0	0	2	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	99.3	100	100
Speaker 'R'	Total errors	6	2	0	2	2	1	0	0	0	3
	Recognition accuracy %	98	99.3	100	99.3	99.3	99	100	100	100	99
Speaker 'W'	Total errors	0	1	0	1	1	0	1	16	1	2
	Recognition accuracy %	100	99.6	100	99.6	99.6	100	99.6	94.7	99.6	99.3

OVERALL TOTAL ERRORS	219
OVERALL RECOGNITION ACCURACY IN %	98.98

SPEAKER-DEPENDENT DIGIT RECOGNITION

		SPOKEN DIGITS									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	1	4	24	6	0	6	3	1	3	0
	Recognition accuracy %	98	92	52	88	100	88	94	98	98	100
Speaker 'A'	Total errors	9	0	8	2	2	3	0	7	0	3
	Recognition accuracy %	82	100	84	96	96	94	100	86	100	94
Speaker 'J'	Total errors	0	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	100
Speaker 'P'	Total errors	2	1	0	0	0	0	0	0	0	3
	Recognition accuracy %	96	98	100	100	100	100	100	100	100	94
Speaker 'E'	Total errors	1	0	0	0	0	0	0	4	0	0
	Recognition accuracy %	98	100	100	100	100	100	100	96	100	100
Speaker 'R'	Total errors	1	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	98	100	100	100	100	100	100	100	100	100
Speaker 'W'	Total errors	5	1	0	5	14	1	0	20	0	0
	Recognition accuracy %	90	98	100	90	72	98	100	60	100	100

OVERALL TOTAL ERRORS	199
OVERALL RECOGNITION ACCURACY IN %	94.32

USING INVERSE FILTER SPECTRAL COEFFICIENTS : 1 to 32 : 'Average distance'

Table 4.10.6

TEXT-INDEPENDENT SPEAKER VERIFICATION

		Spoken digits									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	8	8	2	19	2	106	1	272	0
	Recognition accuracy %	100	97.4	97.4	99.3	93.7	99.3	65.3	99.7	10.8	100
Speaker 'A'	Total errors	0	4	2	0	4	4	0	0	5	0
	Recognition accuracy %	100	98.7	99.3	100	98.7	98.7	100	100	98.4	100
Speaker 'J'	Total errors	0	0	0	0	0	1	0	4	0	0
	Recognition accuracy %	100	100	100	100	100	99.7	100	98.7	100	100
Speaker 'P'	Total errors	4	3	0	0	3	4	1	1	5	37
	Recognition accuracy %	98.7	99	100	100	99	98.7	99.7	99.7	98.4	87.9
Speaker 'E'	Total errors	0	0	0	2	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	99.3	100	100	100	100	100	100
Speaker 'R'	Total errors	0	1	0	1	2	1	0	2	0	9
	Recognition accuracy %	100	99.7	100	99.7	99.3	99.7	100	99.3	100	97
Speaker 'W'	Total errors	0	0	1	1	1	0	2	5	0	1
	Recognition accuracy %	100	100	99.7	99.7	99.7	100	99.3	98.4	100	99.7

OVERALL TOTAL ERRORS	529
OVERALL RECOGNITION ACCURACY IN %	97.52

SPEAKER-DEPENDENT DIGIT RECOGNITION

		Spoken digits									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	3	1	6	6	1	22	0	45	0
	Recognition accuracy %	100	94	88	88.1	88.1	98	56.1	100	10	100
Speaker 'A'	Total errors	0	2	44	0	1	9	0	0	0	0
	Recognition accuracy %	100	96	12	100	98	82	100	100	100	100
Speaker 'J'	Total errors	0	0	0	0	0	14	0	4	0	0
	Recognition accuracy %	100	100	100	100	100	72	100	92	100	100
Speaker 'P'	Total errors	10	0	0	0	0	0	0	0	0	29
	Recognition accuracy %	80	100	100	100	100	100	100	100	100	42
Speaker 'E'	Total errors	0	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	100
Speaker 'R'	Total errors	0	0	0	1	0	0	0	0	0	4
	Recognition accuracy %	100	100	100	98	100	100	100	100	100	92
Speaker 'W'	Total errors	2	0	0	2	12	0	0	10	0	1
	Recognition accuracy %	98	100	100	96	76	100	100	80	100	98

OVERALL TOTAL ERRORS	228
OVERALL RECOGNITION ACCURACY IN %	93.5

Table 4.10.7

USING DIRECT FOURIER TRANSFORM OF SPEECH : 1 to 32 : 'Average distance'

TEXT-INDEPENDENT SPEAKER VERIFICATION

		Spoken digits 0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	3	3	1	2	17	0	2	0	0
	Recognition accuracy %	100	99	99	99.7	99.3	94.4	100	99.3	100	100
Speaker 'A'	Total errors	0	0	4	0	4	2	3	0	5	5
	Recognition accuracy %	100	100	98.7	100	98.7	99.3	99	100	98.3	98.3
Speaker 'J'	Total errors	0	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	100
Speaker 'P'	Total errors	0	1	0	0	4	1	2	0	5	5
	Recognition accuracy %	100	99.7	100	100	98.7	99.7	99.3	100	98.3	98.3
Speaker 'E'	Total errors	0	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	100
Speaker 'R'	Total errors	2	0	0	0	1	25	1	0	0	0
	Recognition accuracy %	99.3	100	100	100	99.7	91.8	99.7	100	100	100
Speaker 'Y'	Total errors	1	0	0	6	0	1	2	2	2	2
	Recognition accuracy %	99.7	100	100	98	100	99.7	99.3	99.3	99.3	99.3

OVERALL TOTAL ERRORS	114
OVERALL RECOGNITION ACCURACY IN %	99.4

SPEAKER-DEPENDENT DIGIT RECOGNITION

		Spoken digits 0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	13	1	0	4	0	11	1	0	0	0
	Recognition accuracy %	74	98	100	92	100	78	98	100	100	100
Speaker 'A'	Total errors	0	0	0	0	0	0	0	0	0	12
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	4 76
Speaker 'J'	Total errors	0	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	100
Speaker 'P'	Total errors	0	0	0	0	0	0	0	0	2	1
	Recognition accuracy %	100	100	100	100	100	100	100	100	96	98
Speaker 'E'	Total errors	0	1	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	98	100	100	100	100	100	100	100	100
Speaker 'R'	Total errors	0	0	0	0	0	19	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	62	100	100	100	100
Speaker 'Y'	Total errors	1	0	0	4	0	5	1	0	0	0
	Recognition accuracy %	98	100	100	92	100	90	98	100	100	100

OVERALL TOTAL ERRORS	76
OVERALL RECOGNITION ACCURACY IN %	97.8

USING LINEAR PREDICTION COEFFICIENTS : 9 to 12 : 'Distance of averages'

Table 4.10.8

TEXT-INDEPENDENT SPEAKER VERIFICATION

					SPOKEN DIGITS						
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	2	5	24	2	5	7	6	9	1	4
	Recognition accuracy %	99.3	98.4	92.1	99.3	98.4	97.7	98	97.1	99.7	98.7
Speaker 'A'	Total errors	8	4	34	3	5	12	2	7	5	11
	Recognition accuracy %	97.4	98.7	88.9	99	98.4	96.1	99.3	97.7	98.4	96.4
Speaker 'J'	Total errors	7	3	2	4	3	4	1	0	0	11
	Recognition accuracy %	97.7	99	99.3	98.7	99	98.7	99.7	100	100	96.4
Speaker 'P'	Total errors	12	16	1	2	6	2	8	13	4	9
	Recognition accuracy %	96.1	94.8	99.7	99.3	98	99.3	97.4	95.7	98.7	97.1
Speaker 'E'	Total errors	0	0	1	0	1	5	4	8	2	2
	Recognition accuracy %	100	100	99.7	100	99.7	98.4	98.7	97.4	99.3	99.3
Speaker 'R'	Total errors	8	1	0	4	1	4	1	3	5	19
	Recognition accuracy %	97.4	99.7	100	98.7	99.7	98.7	99.7	99	98.4	93.8
Speaker 'W'	Total errors	1	2	5	6	6	9	3	12	4	17
	Recognition accuracy %	99.7	99.3	98.4	98	98	97.1	99	96.1	98.7	94.4

OVERALL TOTAL ERRORS	404
OVERALL RECOGNITION ACCURACY IN %	98.1

SPEAKER-DEPENDENT DIGIT RECOGNITION

					SPOKEN DIGITS						
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	6	16	26	7	6	10	6	14	11	9
	Recognition accuracy %	88	68	48	86	88	80	88	72	78	72
Speaker 'A'	Total errors	17	5	44	3	9	15	2	14	1	5
	Recognition accuracy %	66	90	12	94	82	70	96	72	98	90
Speaker 'J'	Total errors	0	0	0	4	0	9	0	1	0	7
	Recognition accuracy %	100	100	100	92	100	82	100	98	100	86
Speaker 'P'	Total errors	5	4	1	5	0	2	5	15	1	8
	Recognition accuracy %	90	92	98	90	100	96	90	70	98	84
Speaker 'E'	Total errors	1	0	5	2	0	1	0	8	0	0
	Recognition accuracy %	98	100	90	96	100	98	100	84	100	100
Speaker 'R'	Total errors	4	3	0	10	0	0	8	3	3	5
	Recognition accuracy %	92	94	100	80	100	100	84	94	94	90
Speaker 'W'	Total errors	3	0	9	7	0	3	0	23	0	6
	Recognition accuracy %	94	100	82	86	100	94	100	54	100	88

OVERALL TOTAL ERRORS	397
OVERALL RECOGNITION ACCURACY IN %	88.7

USING LINEAR PREDICTION COEFFICIENTS: 4 to 9 : 'Distance of averages'

Table 4.10.9

TEXT-INDEPENDENT SPEAKER VERIFICATION

		SPOKEN DIGITS									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	4	76	0	0	6	10	3	4	5
	Recognition accuracy %	100	98.7	76.7	100	100	98	96.7	99	98.7	98.4
Speaker 'A'	Total errors	5	13	20	24	21	8	2	0	3	2
	Recognition accuracy %	98.6	95.7	93.4	92.1	93.1	97.4	99.3	100	99	99.3
Speaker 'J'	Total errors	2	5	1	3	3	2	0	6	1	4
	Recognition accuracy %	99.3	98.4	99.7	99	99	99.3	100	98	99.7	98.7
Speaker 'P'	Total errors	15	8	3	4	8	10	5	8	1	2
	Recognition accuracy %	95	97.4	99	98.7	97.4	96.7	98.4	97.4	99.7	99.3
Speaker 'E'	Total errors	0	8	0	2	11	4	1	5	0	6
	Recognition accuracy %	100	97.4	100	99.3	96.4	98.7	99.7	98.4	100	98
Speaker 'R'	Total errors	3	0	0	0	0	17	3	1	3	3
	Recognition accuracy %	99	100	100	100	100	94.4	99	99.7	99	99
Speaker 'V'	Total errors	10	5	7	3	12	9	16	13	3	24
	Recognition accuracy %	96.7	98.4	97.7	99	96.4	97	94.7	95.7	99	92.1

OVERALL TOTAL ERRORS	465
OVERALL RECOGNITION ACCURACY IN %	97.8

SPEAKER-DEPENDENT DIGIT RECOGNITION

		SPOKEN DIGITS									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	18	28	2	1	10	16	1	14	21
	Recognition accuracy %	100	64	44	96	98	80	68	98	72	38
Speaker 'A'	Total errors	9	14	42	6	36	9	2	8	4	12
	Recognition accuracy %	82	72	16	88	28	82	96	84	92	76
Speaker 'J'	Total errors	0	3	0	5	2	7	0	0	0	3
	Recognition accuracy %	100	94	100	90	96	86	100	100	100	94
Speaker 'P'	Total errors	27	12	1	1	11	19	5	3	0	5
	Recognition accuracy %	46	76	98	98	78	62	90	94	100	90
Speaker 'E'	Total errors	0	3	5	1	37	5	6	2	0	1
	Recognition accuracy %	100	94	90	98	26	90	88	98	100	98
Speaker 'R'	Total errors	0	0	0	0	0	8	1	3	0	1
	Recognition accuracy %	100	100	100	100	100	84	98	94	100	98
Speaker 'V'	Total errors	8	8	2	5	8	13	10	8	0	1
	Recognition accuracy %	84	84	96	90	84	74	80	84	100	98

OVERALL TOTAL ERRORS	493
OVERALL RECOGNITION ACCURACY IN %	85.9

USING LINEAR PREDICTION COEFFICIENTS : 1 to 8 : 'Distance of averages'

TEXT-INDEPENDENT SPEAKER VERIFICATION

Table 4.10.10

		SPOKEN DIGITS 0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	5	17	0	4	9	28	6	5	1
	Recognition accuracy %	100	98.4	94.4	100	98.7	97.1	90.8	98	98.4	99.7
Speaker 'A'	Total errors	3	11	57	7	142	21	2	0	3	4
	Recognition accuracy %	99	96.4	81.3	97.7	53.4	93.1	99.3	100	99	98.7
Speaker 'J'	Total errors	0	7	1	2	1	3	0	1	0	6
	Recognition accuracy %	100	97.7	99.7	99.3	99.7	99	100	99.7	100	98
Speaker 'P'	Total errors	7	4	11	1	4	20	3	5	1	4
	Recognition accuracy %	97.7	98.7	96.4	99.7	98.7	93.4	99	98.4	99.7	98.7
Speaker 'E'	Total errors	0	5	2	1	2	8	2	2	1	31
	Recognition accuracy %	100	98.4	99.3	99.7	99.3	97.4	99.3	99.3	99.7	89.8
Speaker 'R'	Total errors	3	0	0	0	0	13	1	0	1	2
	Recognition accuracy %	99	100	100	100	100	95.7	99.7	100	99.7	99.3
Speaker 'W'	Total errors	3	12	13	5	10	5	17	15	2	22
	Recognition accuracy %	99	96	95.7	98.4	96.7	98.4	94.4	95.1	99.3	92.8

OVERALL TOTAL ERRORS	580
OVERALL RECOGNITION ACCURACY IN %	97.3

SPEAKER-DEPENDENT DIGIT RECOGNITION

		SPOKEN DIGITS 0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	19	17	1	2	13	10	1	14	17
	Recognition accuracy %	100	62	66	98	96	74	80	98	72	66
Speaker 'A'	Total errors	1	11	26	8	37	11	0	2	2	13
	Recognition accuracy %	98	78	48	84	26	78	100	96	96	74
Speaker 'J'	Total errors	0	6	0	3	0	10	0	3	0	3
	Recognition accuracy %	100	88	100	94	100	80	100	94	100	94
Speaker 'P'	Total errors	30	8	8	0	16	12	1	2	0	4
	Recognition accuracy %	40	84	84	100	68	76	98	96	100	92
Speaker 'E'	Total errors	3	7	2	5	29	8	1	2	0	4
	Recognition accuracy %	94	86	96	90	42	84	98	96	100	92
Speaker 'R'	Total errors	0	0	0	0	0	3	1	0	1	1
	Recognition accuracy %	100	100	100	100	100	94	98	100	98	98
Speaker 'W'	Total errors	2	20	1	2	4	13	12	6	0	2
	Recognition accuracy %	96	60	98	96	92	74	76	88	100	96

OVERALL TOTAL ERRORS	847
OVERALL RECOGNITION ACCURACY IN %	75.8

USING LINEAR PREDICTION COEFFICIENTS : 1 to 12 : 'Distance of averages'

Table 4.10.11

TEXT-INDEPENDENT SPEAKER VERIFICATION

Speaker		0	1	2	SPOKEN DIGITS 3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	1	18	0	1	0	9	2	1	4
	Recognition accuracy %	100	99.7	94.1	100	99.7	100	97.1	99.3	99.7	98.7
Speaker 'A'	Total errors	1	4	128	5	16	5	1	0	3	0
	Recognition accuracy %	99.7	98.7	57	98.4	94.8	98.4	99.7	100	99	100
Speaker 'J'	Total errors	0	3	0	0	0	3	0	0	0	2
	Recognition accuracy %	100	99	100	100	100	99	100	100	100	99.3
Speaker 'P'	Total errors	3	3	0	0	7	7	1	2	1	11
	Recognition accuracy %	99	99	100	100	97.7	97.7	99.7	99.3	99.7	96.4
Speaker 'E'	Total errors	0	0	0	0	2	3	0	1	1	0
	Recognition accuracy %	100	100	100	100	99.3	99	100	99.7	99.7	100
Speaker 'R'	Total errors	0	0	0	2	0	1	0	0	1	3
	Recognition accuracy %	100	100	100	99.3	100	99.7	100	100	99.7	99
Speaker 'W'	Total errors	0	3	2	0	1	3	6	5	2	6
	Recognition accuracy %	100	99	99.3	100	99.7	99	98	98.4	99.3	98

OVERALL TOTAL ERRORS	286
OVERALL RECOGNITION ACCURACY IN %	98.7

SPEAKER-DEPENDENT DIGIT RECOGNITION

Speaker		0	1	2	SPOKEN DIGITS 3	4	5	6	7	8	9
Speaker 'S'	Total errors	2	16	18	5	0	1	7	1	11	6
	Recognition accuracy %	96	68	64	90	100	98	86	98	78	88
Speaker 'A'	Total errors	3	6	40	0	26	6	0	7	0	3
	Recognition accuracy %	94	88	20	100	48	88	100	86	100	94
Speaker 'J'	Total errors	0	0	0	0	0	5	0	0	0	3
	Recognition accuracy %	100	100	100	100	100	90	100	100	100	94
Speaker 'P'	Total errors	10	3	0	0	0	3	0	1	0	4
	Recognition accuracy %	80	94	100	100	100	94	100	98	100	92
Speaker 'E'	Total errors	0	0	4	1	0	1	0	1	0	0
	Recognition accuracy %	100	100	92	98	100	98	100	98	100	100
Speaker 'R'	Total errors	0	0	0	0	0	0	0	0	0	1
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	98
Speaker 'W'	Total errors	0	4	3	2	0	3	1	5	0	0
	Recognition accuracy %	100	92	94	96	100	94	98	90	100	100

OVERALL TOTAL ERRORS	213
OVERALL RECOGNITION ACCURACY IN %	93.9

USING INVERSE FILTER SPECTRAL COEFFICIENTS : 1 to 32 : 'Distance of averages'

Table 4.10.12

TEXT-INDEPENDENT SPEAKER VERIFICATION

Speaker		SPOKEN DIGITS 0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	5	49	0	4	0	223	4	184	0
	Recognition accuracy %	100	98.4	83.9	100	98.7	100	26.9	98.7	39.7	100
Speaker 'A'	Total errors	0	2	2	0	5	9	3	0	27	0
	Recognition accuracy %	100	99.3	99.3	100	98.4	97.1	98	100	91.2	100
Speaker 'J'	Total errors	2	0	0	1	1	1	1	7	0	5
	Recognition accuracy %	99.3	100	100	99.7	99.7	99.7	99.7	97.7	100	98.4
Speaker 'P'	Total errors	1	0	0	0	2	4	3	2	0	4
	Recognition accuracy %	99.7	100	100	100	99.3	98.7	99	99.3	100	98.7
Speaker 'K'	Total errors	0	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	100
Speaker 'R'	Total errors	0	1	0	0	2	2	0	0	0	3
	Recognition accuracy %	100	99.7	100	100	99.3	99.3	100	100	100	99
Speaker 'V'	Total errors	3	2	0	0	2	1	4	2	0	2
	Recognition accuracy %	99	99.3	100	100	99.3	99.7	98.7	99.3	100	99.3

OVERALL TOTAL ERRORS	575
OVERALL RECOGNITION ACCURACY IN %	97.31

SPEAKER-DEPENDENT DIGIT RECOGNITION

Speaker		SPOKEN DIGITS 0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	0	18	7	1	2	0	38	0	34	0
	Recognition accuracy %	100	64	86	98	96	100	24	100	32	100
Speaker 'A'	Total errors	0	3	3	0	7	13	0	0	1	2
	Recognition accuracy %	100	94	94	100	86	74	100	100	98	96
Speaker 'J'	Total errors	0	0	0	0	0	11	0	1	0	0
	Recognition accuracy %	100	100	100	100	100	78	100	98	100	100
Speaker 'P'	Total errors	1	0	0	0	2	3	0	0	0	13
	Recognition accuracy %	98	100	100	100	96	94	100	100	100	74
Speaker 'K'	Total errors	0	0	0	0	0	0	0	0	0	0
	Recognition accuracy %	100	100	100	100	100	100	100	100	100	100
Speaker 'R'	Total errors	0	1	0	1	0	0	0	0	0	0
	Recognition accuracy %	100	98	100	98	100	100	100	100	100	100
Speaker 'V'	Total errors	2	1	1	4	1	0	6	0	0	0
	Recognition accuracy %	96	98	98	92	98	100	88	100	100	100

OVERALL TOTAL ERRORS	176
OVERALL RECOGNITION ACCURACY IN %	95.0

USING DIRECT FOURIER TRANSFORM OF SPEECH : 1 to 32 'Distance of averages'

Table 4.10.13

TEXT-INDEPENDENT SPEAKER VERIFICATION

		Spoken digits									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	5	14	36	25	44	64	12	46	0	0
	Recognition accuracy %	98.4	95.4	88.2	91.8	85.5	79	96.1	84.9	100	100
Speaker 'A'	Total errors	1	1	5	0	11	20	2	10	4	9
	Recognition accuracy %	99.7	99.7	98.4	100	96.4	95.4	99.3	96.7	98.7	97.1
Speaker 'J'	Total errors	5	6	2	1	2	2	0	7	0	2
	Recognition accuracy %	98.4	98	99.3	99.7	99.3	99.3	100	97.7	100	99.3
Speaker 'P'	Total errors	1	0	3	3	11	0	3	0	3	5
	Recognition accuracy %	99.7	100	99	99	96.4	100	99	100	99	98.4
Speaker 'E'	Total errors	0	0	0	0	0	2	1	3	0	10
	Recognition accuracy %	100	100	100	100	100	99.3	99.7	99	100	96.7
Speaker 'R'	Total errors	0	0	0	0	1	73	1	0	0	0
	Recognition accuracy %	100	100	100	100	99.7	76.1	99.7	100	100	100
Speaker 'W'	Total errors	6	1	0	22	1	1	17	11	5	4
	Recognition accuracy %	98	99.7	100	92.8	99.7	99.7	94.4	96.4	98.4	98.7

OVERALL TOTAL ERRORS	524
OVERALL RECOGNITION ACCURACY IN %	97.6

SPEAKER-DEPENDENT DIGIT RECOGNITION

		Spoken digits									
		0	1	2	3	4	5	6	7	8	9
Speaker 'S'	Total errors	17	19	9	21	21	24	3	1	0	0
	Recognition accuracy %	66	62	82	58	58	52	94	98	100	100
Speaker 'A'	Total errors	5	3	4	0	15	4	0	11	4	35
	Recognition accuracy %	90	94	92	100	70	92	100	78	92	30
Speaker 'J'	Total errors	0	1	0	1	0	6	0	5	0	0
	Recognition accuracy %	100	98	100	98	100	88	100	90	100	100
Speaker 'P'	Total errors	0	2	0	1	1	0	0	0	3	14
	Recognition accuracy %	100	96	100	98	98	100	100	100	94	72
Speaker 'E'	Total errors	1	0	0	1	0	0	1	2	0	5
	Recognition accuracy %	98	100	100	98	100	100	98	96	100	90
Speaker 'R'	Total errors	0	0	0	0	20	37	1	1	0	0
	Recognition accuracy %	100	100	100	100	60	26	98	98	100	100
Speaker 'W'	Total errors	9	0	0	23	1	18	3	1	0	0
	Recognition accuracy %	82	100	100	54	98	64	94	98	100	100

OVERALL TOTAL ERRORS	354
OVERALL RECOGNITION ACCURACY IN %	89.9

Appendix F

TABULATION OF RESULTS (3-DIGIT STRING EXPERIMENTS)

COMBINED TEXT INDEPENDENT SPEAKER VERIFICATION AND

SPEAKER DEPENDENT 3-DIGIT STRING RECOGNITION

LINEAR PREDICTION COEFFICIENTS 9-12

Table 4.11.1

AVERAGE DISTANCE

		3 DIGIT SEQUENCE						
		387	210	777	888	213	877	037
Speaker S	Total errors	2/9	2/20	0/0	1/0	18/28	2/13	3/19
	Recognition accuracy %	99.0/ 74.2	99.0/ 42.8	100/100	99.5/ 100	91.6/ 20	99.0/ 62.8	98.6 45.7
Speaker A	Total errors	0/15	0/15	1/7	0/0	1/26	0/10	0/19
	Recognition accuracy %	100/ 57.1	100/ 57.1	99.5/ 80	100/100	99.5/ 25.7	100/ 71.4	100/ 45.7
Speaker J	Total errors	1/5	1/1	0/0	0/0	3/2	2/4	6/4
	Recognition accuracy %	99.5/ 85.7	99.5/ 97.1	100/100	100/100	98.6/ 94.2	99.0/ 88.5	97.2/ 88.5
Speaker P	Total errors	7/9	2/3	5/0	7/0	10/13	15/12	16/14
	Recognition accuracy %	96.7/ 74.2	99.0/ 91.4	97.6/ 100	96.7/ 100	95.3/ 62.8	93.0/ 65.7	92.5/ 60
Speaker E	Total errors	2/5	0/0	8/1	1/0	1/1	9/7	1/14
	Recognition accuracy %	99.0/ 85.71	100/100	96.2/ 97.1	99.5/ 100	99.5/ 97.1	95.8/ 80	99.5/ 60
Speaker R	Total errors	2/6	4/0	7/3	0/0	4/0	2/11	3/11
	Recognition accuracy %	99.0/ 82.8	98.1/ 100	96.7/ 91.4	100/100	98.1/ 100	99.0/ 68.5	98.6/ 68.5
Speaker W	Total errors	6/15	1/9	7/27	1/0	7/10	33/18	15/22
	Recognition accuracy %	97.2/ 57.1	99.5/ 74.2	96.7/ 22.8	99.5/ 100	96.7/ 71.4	84.6/ 48.5	93.0/ 37.1

OVERALL TOTAL ERRORS	219/408
OVERALL RECOGNITION ACCURACY IN %	97.9/76.2

DISTANCE OF AVERAGES

		3 DIGIT SEQUENCE						
		387	210	777	888	213	877	037
Speaker S	Total errors	5/9	2/11	19/12	1/11	22/27	4/12	0/5
	Recognition accuracy %	97.6/ 74.2	99.0/ 68.5	91.1/ 65.7	99.5/ 68.5	89.7/ 22.8	98.1/ 65.7	100/ 85.7
Speaker A	Total errors	1/19	8/23	3/12	2/1	1/20	6/23	2/25
	Recognition accuracy %	99.5/ 45.7	96.2/ 35.2	98.6/ 65.7	99.0/ 97.1	99.5/ 42.8	97.2/ 34.2	99.0/ 28.5
Speaker J	Total errors	9/8	5/5	0/0	0/0	2/5	0/8	10/11
	Recognition accuracy %	95.8/ 77.1	97.6/ 85.7	100/100	100/100	99.0/ 85.7	100/ 77.1	95.3/ 68.5
Speaker P	Total errors	2/6	3/4	11/10	7/1	12/11	4/4	9/13
	Recognition accuracy %	99.0/ 82.8	98.6/ 88.5	94.8/ 71.4	96.7/ 97.14	94.4/ 68.5	98.1/ 88.5	95.8/ 62.8
Speaker E	Total errors	7/8	0/1	8/4	2/2	3/4	5/7	2/14
	Recognition accuracy %	96.7/ 77.1	100/ 97.1	96.2/ 88.5	99.0/ 94.2	98.6/ 88.5	97.6/ 80	99.0/ 60
Speaker R	Total errors	3/17	6/13	6/7	3/3	2/3	2/14	2/10
	Recognition accuracy %	98.6/ 51.4	97.2/ 62.8	97.2/ 80	98.6/ 91.4	99.0/ 91.4	99.0/ 60	99.0/ 71.4
Speaker W	Total errors	10/19	5/6	13/28	2/0	11/3	16/15	8/15
	Recognition accuracy %	95.3/ 45.7	97.6/ 82.8	93.9/ 20	99.0/ 100	96.8/ 91.43	92.5/ 57.1	96.2/ 57.1

OVERALL TOTAL ERRORS	266/489
OVERALL RECOGNITION ACCURACY IN %	97.4/71.4

TIDSV / SDDR

LINEAR PREDICTION COEFFICIENTS 4-9

Table 4.11.2

AVERAGE DISTANCE

		3 DIGIT SEQUENCE						
		387	210	777	888	213	877	037
Speaker S	Total errors	1/12	0/12	0/0	3/4	7/23	3/13	0/12
	Recognition accuracy %	99.5/ 65.7	100/ 65.7	100/100	98.6/ 88.5	96.7/ 34.2	98.6/ 62.8	100/ 65.7
Speaker A	Total errors	2/12	0/3	1/0	2/0	3/13	1/8	0/4
	Recognition accuracy %	99.0/ 65.7	100/ 91.4	99.5/ 100	99.0/ 100	98.6/ 62.8	99.5/ 77.1	100/ 88.5
Speaker J	Total errors	0/5	0/0	0/0	0/0	1/0	0/4	0/0
	Recognition accuracy %	100/ 85.7	100/100	100/100	100/100	99.5/ 100	100/ 88.5	100/100
Speaker P	Total errors	6/8	7/0	7/3	2/0	6/0	2/9	10/9
	Recognition accuracy %	97.2/ 77.1	96.7/ 100	96.7/ 91.4	99.0/ 100	97.2/ 100	99.0/ 76.2	95.3/ 74.2
Speaker E	Total errors	0/5	0/0	0/0	0/0	1/7	0/6	0/13
	Recognition accuracy %	100/ 85.7	100/100	100/100	100/100	99.5/ 80	100/ 82.8	100/ 62.86
Speaker R	Total errors	0/1	0/0	8/0	1/0	0/0	6/3	1/2
	Recognition accuracy %	100/ 97.1	100/100	96.2/ 100	99.5/ 100	100/100	97.2/ 91.4	99.5/ 94.2
Speaker W	Total errors	10/5	2/3	28/5	2/0	5/4	37/9	2/4
	Recognition accuracy %	95.8/ 85.7	99.0/ 91.4	86.9/ 85.7	99.0/ 100	97.6/ 88.5	82.7/ 74.2	99.0/ 88.5

OVERALL TOTAL ERRORS	167/221
OVERALL RECOGNITION ACCURACY IN %	98.4/87.1

DISTANCE OF AVERAGES

		3 DIGIT SEQUENCE						
		387	210	777	888	213	877	037
Speaker S	Total errors	2/7	0/16	6/0	3/16	5/28	3/11	0/7
	Recognition accuracy %	99.0/ 80	100/ 54.2	97.2/ 100	98.6/ 54.2	97.6 20	98.6/ 68.5	100/ 80
Speaker A	Total errors	0/9	0/5	0/0	2/13	0/4	0/3	0/22
	Recognition accuracy %	100/ 74.2	100/ 85.7	100/ 100	99.0/ 62.8	100/ 88.5	100/ 91.4	100/ 37.1
Speaker J	Total errors	3/4	2/4	6/1	1/0	3/9	0/5	7/13
	Recognition accuracy %	98.6/ 88.5	99.0/ 88.5	97.2/ 97.1	99.5/ 100	98.6/ 74.2	100/ 85.7	96.7/ 62.8
Speaker P	Total errors	4/1	6/5	12/12	1/0	15/11	11/8	12/18
	Recognition accuracy %	98.1/ 97.1	97.2/ 85.7	94.4/ 65.7	99.5/ 100	93.0/ 68.5	94.8/ 77.1	94.4/ 48.5
Speaker E	Total errors	0/2	0/5	3/5	0/1	0/8	1/7	0/2
	Recognition accuracy %	100/ 94.2	100/ 85.7	98.6/ 85.7	100/ 97.1	100/ 77.1	99.5/ 80	100/ 94.2
Speaker R	Total errors	0/3	0/0	1/3	0/0	2/1	0/11	1/9
	Recognition accuracy %	100/ 91.4	100/ 100	99.5/ 91.4	100/ 100	99.0/ 97.1	100/ 68.5	99.5/ 74.2
Speaker W	Total errors	7/8	5/6	14/10	7/1	22/19	24/8	6/17
	Recognition accuracy %	96.7/ 77.1	97.6/ 82.8	93.4/ 71.4	96.7/ 97.1	89.7/ 45.7	88.8/ 77.1	97.2/ 51.4

OVERALL TOTAL ERRORS	197/381
OVERALL RECOGNITION ACCURACY IN %	98.1/77.7

TIDSV / SDDR

Table 4.11.3

LINEAR PREDICTION COEFFICIENTS 1-8

		AVERAGE DISTANCE: 3 DIGIT SEQUENCE							DISTANCE OF AVERAGES: 3 DIGIT SEQUENCE						
		387	210	777	888	213	877	037	387	210	777	888	213	877	037
Speaker S	Total errors	1/8	0/3	3/0	3/8	0/8	2/11	0/1	1/6	0/18	6/0	2/15	2/21	5/10	0/5
	Recognition accuracy %	99.5/ 77.1	100/ 91.4	98.6/ 100	98.6/ 77.1	100/ 77.1	99.0/ 68.5	100/ 97.1	99.5/ 82.8	100/ 48.5	97.2/ 100	99.0/ 57.1	99.0/ 40	97.6/ 71.4	100/ 85.7
Speaker A	Total errors	2/8	1/2	0/0	2/0	2/4	0/5	0/0	0/4	0/10	0/0	3/12	1/10	0/5	0/20
	Recognition accuracy %	99.0/ 77.1	99.5/ 94.2	100/100	99.0/ 100	99.0/ 88.5	100/ 85.7	100/100	100/ 88.5	100/ 71.4	100/ 100	98.6/ 65.7	99.5/ 71.4	100/ 85.7	100/ 42.8
Speaker J	Total errors	0/0	0/0	2/0	1/0	0/0	0/0	0/0	4/6	3/3	3/0	0/0	3/20	1/0	1/9
	Recognition accuracy %	100/100	100/100	99.0/ 100	99.5/ 100	100/100	100/100	100/100	98.1/ 82.8	98.6/ 91.4	98.6/ 100	100/ 100	98.6/ 42.8	99.5/ 100	99.5/ 74.2
Speaker P	Total errors	1/3	0/0	4/3	9/0	0/0	2/5	1/1	9/0	1/6	4/8	1/0	0/2	19/3	9/22
	Recognition accuracy %	99.5/ 91.4	100/100	98.1/ 91.4	95.8/ 100	100/100	99.0/ 85.7	99.5/ 97.1	95.8/ 100	99.5/ 82.86	98.1/ 77.1	99.5/ 100	100/ 94.2	91.1/ 91.4	95/8/ 37.1
Speaker E	Total errors	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	2/5	3/5	1/0	0/8	0/1	0/2
	Recognition accuracy %	100/100	100/100	100/100	100/100	100/100	100/100	100/100	100/ 100	99.0/ 85.7	98.6/ 85.7	99.5/ 100	100/ 77.1	100/ 97.1	100/ 94.2
Speaker R	Total errors	0/0	0/0	0/0	1/0	0/0	0/1	0/0	0/0	0/0	0/1	0/0	0/0	0/1	1/2
	Recognition accuracy %	100/100	100/100	100/100	99.5/ 100	100/100	100/ 97.1	100/100	100/ 100	100/ 100	100/ 97.1	100/ 100	100/ 100	100/ 97.1	99.5/ 94.2
Speaker V	Total errors	2/3	1/1	16/0	0/0	0/0	7/5	0/0	5/1	8/7	12/8	3/0	12/18	9/3	1/11
	Recognition accuracy %	99.0/ 91.4	99.5/ 97.1	92.5/ 100	100/100	100/100	96.7/ 85.7	100/100	97.6/ 97.1	96.2/ 80	94.4/ 77.1	98.6/ 100	96.6/ 48.5	95.8/ 91.4	99.5/ 68.5

	AVERAGE DISTANCE	DISTANCE OF AVERAGES
OVERALL TOTAL ERRORS	63/80	135/288
OVERALL RECOGNITION ACCURACY IN %	99.4/95.3	98.7/83.2

TIDSV / SDDR

LINEAR PREDICTION COEFFICIENTS 1-12

Table 4.11.4

		AVERAGE DISTANCE: 3 DIGIT SEQUENCE							DISTANCE OF AVERAGES: 3 DIGIT SEQUENCE						
		387	210	777	888	213	877	037	387	210	777	888	213	877	037
Speaker S	Total errors	1/9	0/15	0/0	2/2	14/26	1/12	0/14	0/4	0/13	1/0	1/14	0/21	0/7	0/6
	Recognition accuracy %	99.5/ 74.2	100/ 57.1	100/100	99.0/ 94.2	93.4/ 25.7	99.5/ 65.7	100/ 60	100/ 88.5	100/ 62.8	99.5/ 100	99.5/ 60	100/ 40	100/ 80	100/ 82.8
Speaker [illegible]	Total errors	0/9	0/3	0/0	1/0	1/6	0/4	0/2	0/2	2/9	0/3	3/4	1/8	0/4	0/19
	Recognition accuracy %	100/ 74.2	100/ 91.4	100/100	99.5/ 100	99.5/ 82.8	100/ 88.5	100/ 94.2	100/ 94.2	99.0/ 74.2	100/ 91.4	98.6/ 88.5	99.5/ 77.1	100/ 88.5	100/ 45.7
Speaker J	Total errors	0/0	0/0	0/0	0/0	0/0	0/0	0/0	2/3	0/2	0/0	0/0	0/2	0/0	2/6
	Recognition accuracy %	100/100	100/100	100/100	100/100	100/100	100/100	100/100	99.0/ 91.4	100/ 94.2	100/100	100/100	100/ 94.2	100/100	99.0/ 82.8
Speaker P	Total errors	1/4	1/0	2/0	6/0	0/0	3/8	0/0	2/0	1/3	1/4	1/0	2/3	2/2	5/19
	Recognition accuracy %	99.5/ 88.5	99.5/ 100	99.0/ 100	97.2/ 100	100/100	98.6/ 77.1	100/100	99.0/ 100	99.5/ 91.4	99.5/ 88.5	99.5/ 100	99.0/ 91.4	99.0/ 94.2	97.6/ 45.7
Speaker E	Total errors	0/4	0/0	0/0	0/0	0/0	0/3	0/0	0/1	0/1	1/4	1/1	0/1	0/3	0/1
	Recognition accuracy %	100/ 88.5	100/100	100/100	100/100	100/100	100/ 91.4	100/100	100/ 97.1	100/ 97.1	99.5/ 88.5	99.5/ 97.1	100/ 97.1	100/ 91.4	100/ 97.1
Speaker R	Total errors	0/1	0/0	0/0	0/0	0/0	0/0	0/0	1/12	0/0	0/0	0/1	0/1	0/3	0/6
	Recognition accuracy %	100/ 97.1	100/100	100/100	100/100	100/100	100/100	100/100	99.5/ 65.7	100/100	100/100	100/ 97.1	100/ 97.1	100/ 94.2	99.5/ 82.8
Speaker W	Total errors	6/5	0/3	13/5	1/0	0/2	13/6	2/5	3/1	0/6	10/9	2/0	2/4	10/2	0/4
	Recognition accuracy %	97.2/ 85.7	100/ 91.4	93.9/ 85.7	99.5/ 100	100/ 94.2	93.9/ 82.8	99.0/ 85.7	98.6/ 97.1	100/ 82.8	95.3/ 74.2	99.0/ 100	99.0/ 88.5	95.3/ 94.2	100/ 88.5

AVERAGE DISTANCE:

OVERALL TOTAL ERRORS	68/148
OVERALL RECOGNITION ACCURACY IN %	99.3/91.3

DISTANCE OF AVERAGES:

OVERALL TOTAL ERRORS	57/218
OVERALL RECOGNITION ACCURACY IN %	99.4/87.2

TIDSV / SDDR

DIRECT FOURIER TRANSFORM OF SPEECH 1-32

Table 4.11.5

		AVERAGE DISTANCE: 3 DIGIT SEQUENCE 210	777	888	DISTANCE OF AVERAGES: 3 DIGIT SEQUENCE 210	777	888
Speaker S	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 1 / 5	24 / 25 / 0	0 / 0 / 0
	Recognition accuracy %	100/100/100	100/100/100	100/100/100	100/98.9/66.7	31.4/73.7/100	100/100/100
Speaker A	Total errors	0 / 0 / 0	0 / 0 / 0	2 / 2 / 0	0 / 0 / 0	4 / 4 / 1	4 / 4 / 0
	Recognition accuracy %	100/100/100	100/100/100	96 / 98 /100	100/100/100	88.6/95.8/93.3	88.6/95.8/100
Speaker J	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	3 / 3 / 0	4 / 4 / 0	0 / 0 / 0
	Recognition accuracy %	100/100/100	100/100/100	100/100/100	91.4/96.8/100	88.6/95.8/100	100/100/100
Speaker P	Total errors	0 / 0 / 0	0 / 0 / 0	6 / 6 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	Recognition accuracy %	100/100/100	100/100/100	84 /94 /100	100/100/100	100/100/100	100/100/100
Speaker E	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	Recognition accuracy %	100/100/100	100/100/100	100/100/100	100/100/100	100/100/100	100/100/100
Speaker R	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	Recognition accuracy %	100/100/100	100/100/100	100/100/100	100/100/100	100/100/100	100/100/100
Speaker W	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	1 / 1 / 0	12/12/0	0 / 0 / 0
	Recognition accuracy %	100/100/100	100/100/100	100/100/100	97.1/98.9/100	65.7/87.4/100	100/100/100

	AVERAGE DISTANCE	DISTANCE OF AVERAGES
OVERALL TOTAL ERRORS	8 / 8 / 0	52 / 54 / 6
OVERALL RECOGNITION ACCURACY IN %	98.9/99.6/100	92.9/97.3/98.1

TDSV / TIDSV / SDDR

Appendix G

TABULATION OF RESULTS (7-DIGIT STRING EXPERIMENTS)

COMBINED TEXT INDEPENDENT SPEAKER VERIFICATION AND

SPEAKER DEPENDENT 7-DIGIT STRING RECOGNITION

LINEAR PREDICTION COEFFICIENTS 9-12

Table 4.12.1

		AVERAGE DISTANCE: 7 DIGIT SEQUENCE '2017839'	AVERAGE DISTANCE: 7 DIGIT SEQUENCE '7926843'	DISTANCE OF AVERAGES: 7 DIGIT SEQUENCE '2017839'	DISTANCE OF AVERAGES: 7 DIGIT SEQUENCE '7926843'
Speaker 'S'	Total errors	0 / 0 / 0	0 / 0 / 3	2 / 2 / 3	4 / 3 / 5
	Recognition accuracy %	100/100/100	100/100/70	94.3/96.9/70	88.6/95.4/50
Speaker 'A'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 3	0 / 0 / 5
	Recognition accuracy %	100/100/100	100/100/100	100/100/70	100/100/50
Speaker 'J'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 2	1 / 1 / 5
	Recognition accuracy %	100/100/100	100/100/100	100/100/80	97.1/98.5/50
Speaker 'P'	Total errors	2 / 2 / 0	4 / 5 / 4	7 /10/ 4	4 / 7 / 5
	Recognition accuracy %	94.3/96.9/100	88.6/92.3/60	80/87.7/60	88.6/89.2/50
Speaker 'E'	Total errors	0 / 0 / 0	0 / 0 / 0	2 / 2 / 5	0 / 0 / 4
	Recognition accuracy %	100/100/100	100/100/100	94.3/96.9/50	100/100/60
Speaker 'R'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 3	0 / 0 / 4
	Recognition accuracy %	100/100/100	100/100/100	100/100/70	100/100/60
Speaker 'V'	Total errors	11/16/4	26/45/5	11/27/2	26/39/5
	Recognition accuracy %	68.6/75.4/60	25.7/30.7/50	68.6/58.5/80	25.7/40/50

AVERAGE DISTANCE		DISTANCE OF AVERAGES	
OVERALL TOTAL ERRORS	43/68/16	OVERALL TOTAL ERRORS	57/91/55
OVERALL RECOGNITION ACCURACY IN %	91.2/92.5/88.6	OVERALL RECOGNITION ACCURACY IN %	88.4/91/60.7

TDSV / TIDSV / SDDR

LINEAR PREDICTION COEFFICIENTS 4-9

Table 4.12.2

		AVERAGE DISTANCE: 7 DIGIT SEQUENCE '2017839'	AVERAGE DISTANCE: 7 DIGIT SEQUENCE '7926843'	DISTANCE OF AVERAGES: 7 DIGIT SEQUENCE '2017839'	DISTANCE OF AVERAGES: 7 DIGIT SEQUENCE '7926843'
Speaker 'S'	Total errors	0 / 0 / 0	0 / 0 / 3	0 / 0 / 0	0 / 0 / 5
	Recognition accuracy %	100/100/100	100/100/70	100/100/100	100/100/50
Speaker 'A'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 5	0 / 0 / 4
	Recognition accuracy %	100/100/100	100/100/100	100/100/50	100/100/60
Speaker 'J'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 1	0 / 0 / 5
	Recognition accuracy %	100/100/100	100/100/100	100/100/90	100/100/50
Speaker 'P'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 4	0 / 0 / 3
	Recognition accuracy %	100/100/100	100/100/100	100/100/60	100/100/70
Speaker 'E'	Total errors	0 / 0 / 0	0 / 0 / 0	3 / 3 / 5	1 / 2 / 4
	Recognition accuracy %	100/100/100	100/100/100	91.4/95.4/50	97.1/96.9/60
Speaker 'R'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 3	0 / 3 / 3
	Recognition accuracy %	100/100/100	100/100/100	100/100/70	100/95.4/70
Speaker 'V'	Total errors	0 / 0 / 0	0 / 0 / 0	10/19/5	10/16/3
	Recognition accuracy %	100/100/100	100/100/100	71.4/70.8/50	71.4/75.4/70

AVERAGE DISTANCE	
OVERALL TOTAL ERRORS	0 / 0 / 3
OVERALL RECOGNITION ACCURACY IN %	100/100/97.9

DISTANCE OF AVERAGES	
OVERALL TOTAL ERRORS	24/43/50
OVERALL RECOGNITION ACCURACY IN %	95.1/95.3/64.3

TDSV / TIDSV / SDDR

LINEAR PREDICTION COEFFICIENTS 1-8

Table 4.12.3

Speaker		Average Distance, 7 digit sequence: '2017839'	Average Distance, 7 digit sequence: '7926843'	Distance of Averages, 7 digit sequence: '2017839'	Distance of Averages, 7 digit sequence: '7926843'
Speaker 'S'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 1	0 / 0 / 5
	Recognition accuracy %	100/100/100	100/100/100	100/100/90	100/100/50
Speaker 'A'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 3	0 / 0 / 5
	Recognition accuracy %	100/100/100	100/100/100	100/100/70	100/100/50
Speaker 'J'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 1	0 / 0 / 1
	Recognition accuracy %	100/100/100	100/100/100	100/100/90	100/100/90
Speaker 'P'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 3	0 / 0 / 1
	Recognition accuracy %	100/100/100	100/100/100	100/100/70	100/100/90
Speaker 'E'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 5	0 / 0 / 4
	Recognition accuracy %	100/100/100	100/100/100	100/100/50	100/100/60
Speaker 'R'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 1 / 3
	Recognition accuracy %	100/100/100	100/100/100	100/100/100	100/98.5/70
Speaker 'W'	Total errors	0 / 0 / 0	0 / 0 / 0	3 / 3 / 0	6 / 9 / 5
	Recognition accuracy %	100/100/100	100/100/100	91.4/95.4/100	82.9/86.1/50
OVERALL TOTAL ERRORS		0 / 0 / 0		9/13/37	
OVERALL RECOGNITION ACCURACY IN %		100/100/100		98.1/98.6/73.6	

TDSV / TIDSV / SDDR

LINEAR PREDICTION COEFFICIENTS 1-12

Table 4.12.4

		AVERAGE DISTANCE: 7 DIGIT SEQUENCE '2017839'	AVERAGE DISTANCE: 7 DIGIT SEQUENCE '7926843'	DISTANCE OF AVERAGES: 7 DIGIT SEQUENCE '2017839'	DISTANCE OF AVERAGES: 7 DIGIT SEQUENCE '7926843'
Speaker 'S'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	Recognition accuracy %	100/100/100	100/100/100	100/100/100	100/100/100
Speaker 'A'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 2	0 / 0 / 5
	Recognition accuracy %	100/100/100	100/100/100	100/100/80	100/100/50
Speaker 'J'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 1	0 / 0 / 5
	Recognition accuracy %	100/100/100	100/100/100	100/100/90	100/100/50
Speaker 'P'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 1 / 4	1 / 1 / 5
	Recognition accuracy %	100/100/100	100/100/100	100/98.5/60	97.1/98.5/50
Speaker 'E'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 5	0 / 0 / 4
	Recognition accuracy %	100/100/100	100/100/100	100/100/50	100/100/60
Speaker 'R'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 1	0 / 0 / 4
	Recognition accuracy %	100/100/100	100/100/100	100/100/90	100/100/60
Speaker 'N'	Total errors	5 / 5 / 3	7 / 7 / 4	9 / 13 / 1	12/22/5
	Recognition accuracy %	85.7/92.3/70	80/89.2/60	74.3/ 80 /90	65.8/66.2/50

	AVERAGE DISTANCE		DISTANCE OF AVERAGES
OVERALL TOTAL ERRORS	12/12/7	OVERALL TOTAL ERRORS	22/37/47
OVERALL RECOGNITION ACCURACY IN %	97.6/98.7/95	OVERALL RECOGNITION ACCURACY IN %	95.5/95.9/66.4

TDSV / TIDSV / SDDR

Table 4.12.5

DIRECT FOURIER TRANSFORM OF SPEECH 1-32

		AVERAGE DISTANCE: 7 DIGIT SEQUENCE '2017839'	AVERAGE DISTANCE: 7 DIGIT SEQUENCE '7926843'	DISTANCE OF AVERAGES: 7 DIGIT SEQUENCE '2017839'	DISTANCE OF AVERAGES: 7 DIGIT SEQUENCE '7926843'
Speaker 'A'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 1 / 5	0 / 0 / 3
	Recognition accuracy %	100/100/100	100/100/100	100/97.78/50	100/100/70
Speaker 'J'	Total errors	0 / 0 / 0	0 / 0 / 0	1 / 2 / 4	1 / 1 / 3
	Recognition accuracy %	100/100/100	100/100/100	96/95.56/60	96/97.78/70
Speaker 'E'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 4	0 / 0 / 3
	Recognition accuracy %	100/100/100	100/100/100	100/100/60	100/100/70
Speaker 'R'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
	Recognition accuracy %	100/100/100	100/100/100	100/100/100	100/100/100
Speaker 'S'	Total errors	0 / 0 / 0	0 / 0 / 0	0 / 0 / 5	0 / 0 / 3
	Recognition accuracy %	100/100/100	100/100/100	100/100/50	100/100/70

TDSV/TIDSV/SDDR

	AVERAGE DISTANCE	DISTANCE OF AVERAGES
OVERALL TOTAL ERRORS	0 / 0 / 0	2 / 4 / 30
OVERALL RECOGNITION ACCURACY IN %	100/100/100	99.2/99.1/70

TDSV / TIDSV / SDDR

LIST OF REFERENCES

1. Alphonse Chapanis 'Interactive Human Communications' Scientific American, vol 232, No.3, 1975, pp 36-42

2. A.E.Rosenberg and K.L.Shipley 'Speaker Identification and Verification Combined with Speaker-Independent Word Recognition' IEEE Trans., in Acoust., Speech & Signal Processing, 1981, pp 184-187.

3. M.R.Sambur and L.R.Rabiner 'A Speaker-Independent Digit Recognition System' The Bell Systems Technical Journal, Jan., 1975, pp 151-172

4. Fumitada Itakura 'Minimum Prediction Residual Principle Applied to Speech Recognition' IEEE Trans., Acoust., Speech and Signal Processing, vol ASSP-23, Feb., 1975, pp 67-72

5. Marvin R. Sambur 'Speaker Recognition Using Orthogonal Linear Prediction' IEEE Trans., Acoust., Speech and Signal Processing, vol ASSP-24, Aug., 1976, pp 283-289.

6. N.Mohankrishnan 'Text-Independent Speaker Recognition' Ph.D thesis, Dept., of Electl., Engg., Univ., of Windsor, Canada, 1984.

7. L.R.Rabiner, A.E.Rosenberg, Stephen E.Levinson 'Considerations in Dynamic Time Warping Algorithms for Discrete Word Recognition' IEEE Trans., on Acoust., Speech and Signal Processing, vol ASSP-26, No.6, Dec., 1978, pp 575-581

8. Hiroaki Sakoe, Seibi Chiba 'Dynamic Programming Algorithm Optimization for Spoken Word Recognition' IEEE Trans., on Acoust., Speech and Signal Processing, vol ASSP-26, No.1, Feb., 1978.

9. J.D.Markel 'Linear Prediction of Speech' Springer-Verlag Berlin Heidelberg, New York, 1976.

10. Wakita 'Direct Estimation of The Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms' IEEE Trans., on Audio Electro Acoustics, vol AU-21, Oct., 1973, pp 417 427.

11. L.R.Rabiner,M.R.Sambur 'An Algorithm for Determining the Endpoints of Isolated Utterances' Bell Systems Technical Journal, 54, No.2,Feb.,1975.

12. F.J.Harris 'On the Use of Windows for Harmonic Analysis with the DFT' Proceedings of IEEE, January 1978.

13. B.S.Atal 'Effectiveness of Linear Prediction Characteristics of the Speech Wave for Automatic Speaker Identification And Verification' Journal of Acoust., Society of America,vol.55,No.6,June 1974,pp 1304- 1312.

14. Stephen S.McCandles 'An Algorithm for Automatic Formant Extraction Using Linear Prediction Spectra' IEEE Trans., on Acoust., Speech and Signal Processing,April1974,pp 312- 318.

15. Robert E.Bogner 'On Talker Verification via Orthogonal Parameters' IEEE Trans., on Acoust., Speech and Signal Processing,vol ASSP-29,No.1, February 1981,pp 1-12.

16. A.E.Rosenberg 'Automatic Speaker Verification : A Review' Proceedings of the IEEE,vol 64, April 1976, pp 475-487.

17. Bishnu S.Atal 'Automatic Recognition of Speakers from Their Voices' Proceedings of the IEEE, vol 64, April 1976, pp 460-475.

18. S.K.Das,W.S.Mohn 'A Scheme for Speech Processing in Automatic Speaker Verification' IEEE Trans., on Audio and Electro Acoustics, vol AU-19, March 1971, pp 32-43.

19. George R.Doddington 'Personal Identity Verification Using Voice' Proceedings Electro-76, May 11-14, pp 22-4,1-5.

20. L.R.Rabiner,S.E.Levinson,A.E.Rosenberg,J.G.Wilpon 'Speaker-Independent Recognition of Isolated Words Using Clustering Techniques' IEEE Trans., on Acoust., Speech and Signal Processing,vol ASSP-27,No.4, August 1979, pp 336-349.

21. L.R.Rabiner,R.W.Schafer 'Digital Processing of Speech Signals' Prentice-Hall Inc.,N.J.

22. J.L.Flanagan 'Computers That Talk and Listen: Man-Machine Communication by voice' Proceedings of the IEEE,vol 64,April 1976,pp 405-415.

23. T.B.Martin 'Practical Appliactions of Voice Input to Machines' Proceedings of the IEEE,vol 64,April 1976,pp 487-501.

24. Wayne A.Lea 'Trends in Speech Recognition' Prentice-Hall Signal Processing Series, N.J.,1980

25. Stephen E.Levinson,Mark Y.Liberman 'Speech Recognition by Computer' Scientific American, June 1981, pp 64-76.

26. Robert C.Lummis 'Speaker Verification by Computer Using Speech Intensity for Temporal Registration' IEEE Trans., on Audio Electro Acoustics, vol AU-21,No.2,April 1973, pp 80-88.

27. James E.Luck 'Automatic Speaker Verification Using Cepstral Measurements' The Journal of Acoustical Society of America,vol 46,No.4,April 1969, pp 1026-1030.

28. Thomas B.Schalk,Elizabeth L.Van Meir 'Terminals, Listen Up, Speech Recognition Is A Reality' Computer Design, September 1983, pp 97-102.

29. L.R.Rabiner,J.G.Wilpon 'Considerations in Applying Clustering Techniques to Speaker-Independent Word Recognition' Journal of Acoustic Society of America, 66(3), Sept.,1979, pp 663-673.

30. Stephen E.Levinson,L.R.Rabiner,A.E.Rosenberg,J.G.Wilpon 'Interactive Clustering Techniques for Selecting Speaker-Independent Reference Templates for Isolated Word Recognition' IEEE Trans., on Acoust., Speech and Signal Processing,vol ASSP-27,No.2, April 1979,pp 134-140.

31. M.Shridhar 'A Unified Approach to Speaker Verification with Noisy Speech Inputs' Speech Communication 1 (1982), pp 103-112.

32. M.Shridhar,N.Mohankrishnan 'Text-Independent Speaker Recognition : A Review And Some New Results' Speech Communication 1 (1982), pp 257-267.

33. Bruce A. Sherwood, University of Illinois. 'The computer speaks' IEEE Spectrum vol.16, No. 8, August 1979, pp. 18-25.

34. Thomas B. Martin, Threshold Technology. 'One way to talk to computers' IEEE Spectrum, vol.14, No.5, May 1977, pp. 35-39.

35. Gadi Kaplan, Raj Reddy, Yasuo Kato 'Words into action' IEEE Spectrum, vol 17, No. 6, June 1980, pp. 22-33.

VITA AUCTORIS

1952	Born in Kasargod, India.
1968	Graduated from St. Aloysius High School, Mangalore, India, with S.S.L.C in Science.
1969	Graduated from St.Aloysius College Mangalore, India, with Pre University Course, in Science.
1974	Graduated from Karnataka Regional Engineering College, Surathkal, Mangalore, India, with Bachelor of Engineering, in Electronics and Communications.
1974-1976	Employed in International Meters & Electronics Corporation (IMECO Ultrasonics Pvt. Ltd.), Bombay, as Electronics design engineer. Engaged in analog and digital circuits design of their medical and industrial products.
1976-1979	Employed in Institute for Design of Electrical Measuring Instruments (IDEMI), Bombay, as Field Officer (Electronics). Engaged in design and development of process control instrumentation.
1979	Graduated from Jamnalal Bajaj Institute of Management Studies, Bombay, with Diploma in Computer Management.
1979-1982	Employed in Tata Institute of Fundamental Research, Bombay, as Scientific Officer. Engaged in the fields of multiplexed automatic voice answering systems; voice, telex, computer data switching system; bit-slice architectures; hardware of DEC-PDP compatible computers.
1982-1984	Candidate for Master of Applied Science degree, Electrical Engg., University of Windsor, Windsor, Ontario, Canada.