

Word recognition from speech signal using Spectrum Analysis and LPC

*A thesis submitted in partial fulfillment of the requirements for the award of
degree of*

Master of Engineering

in

Electronic Instrumentation and Control



Submitted by
Gurpreet Singh
Roll No. 800951011

Under the Guidance of
Mr. M.D Singh
Assistant Professor

Department of Electrical and Instrumentation Engineering
Thapar University

(Established under the section 3 of UGC act, 1956)

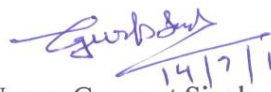
Patiala, 147004, Punjab, India

July 2011

CERTIFICATE

I hereby certify that the work which is being presented in the thesis entitled “**Word recognition from speech signal using Spectrum Analysis and LPC**” in partial fulfillment of award of degree of **Master of Engineering in Electronics Instrumentation and Control** submitted in Electrical and Instrumentation Engineering department, Thapar University, Patiala is an authentic record of my own work carried under the supervision of **Mr. M.D Singh**, Assistant Professor, Department of Electrical and Instrumentation Engineering, Thapar University, Patiala, Punjab.

Date: 14.7.11

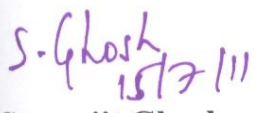

Name: Gurpreet Singh
Roll No: 800951011

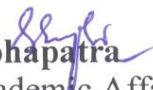
I certify that the above statement made by the student is correct to the best of my knowledge and belief.

Date: 14/7/11


Mr. M.D Singh
Assistant Professor
Department of Electrical and
Instrumentation Engineering
Thapar University, Patiala
Punjab

Countersigned By


Dr. Smarajit Ghosh
Head of Department
Department of Electrical and
Instrumentation Engineering
Thapar University, Patiala
Punjab


Dr. S K Mohapatra
Dean of Academic Affairs
Thapar University, Patiala
Punjab

ABSTRACT

In this thesis, word recognition has been done from the speech signal using spectrum analysis. Word recognition involves features extraction, preprocessing, pattern matching and decision making. To recognize the word we find word parameters from spectrum of speech signal. Here, we collect five samples of a single word and then analysis on their spectrum. To find the word parameters we analysis the word spectrum by statistical methods which are given to the some range values. These parameter values help us to recognition the word. From these values we shown that every word parameters has their own value and every parameter lying in between some bounded or range of values. Out of these range values the parameter of word has changed. Parameters finding of speech signal is the key factor to improve the recognition word. To find these parameters we use some statistical methods. And find another different parameters we use the other method which is Linear Predicted Coding (LPC) coefficients. From this method we get the new signal spectrum which are made and different from the original speech signal spectrum. So by this new spectrum analysis we find another new parameters which are help us to recognition the word. This is developed in MATLAB. So in this work, we collect the twenty five words each have five samples and analysis their signal spectrum and find various parameters. From these parameters values we recognition the word.

ACKNOWLEDGEMENT

The real spirit of achieving a goal is through the way of excellence and austere discipline. I would have never succeeded in completing my task without the cooperation, encouragement and help provided to me by various personalities.

With deep sense of gratitude I express my sincere thanks to my esteemed and worthy supervisor, **Mr. M. D. SINGH** , Assistant Professor, Department of Electrical and Instrumentation Engineering, Thapar University, Patiala for his valuable guidance in carrying out this work under his effective supervision, encouragement, enlightenment and cooperation. Most of the novel ideas and solutions found in this thesis are the result of our numerous stimulating discussions. His feedback and editorial comments were also invaluable for writing of this thesis.

I shall be failing in my duties if I do not express my deep sense of gratitude towards **Dr. Smarajit Ghosh**, Professor and Head of the Department of Electrical & Instrumentation Engineering, Thapar University, Patiala who has been a constant source of inspiration for me throughout this work.

I am also thankful to all the staff members of the Department for their full cooperation and help.

This acknowledgement would be incomplete if I do not mention the emotional support and blessings provided by my friends. I had a pleasant enjoyable and fruitful company with them.

My greatest thanks are to all who wished me success especially my parents, sister whose support and care makes me stay on earth.

Place: Thapar University, Patiala
Date:

Gurpreet Singh
800951011

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE NO.
	CERTIFICATE	i
	ABSTRACT	ii
	ACKNOWLEDGEMENT	iii
	TABLE OF CONTENTS	iv - v
	LIST OF FIGURES	vi - vii
	LIST OF TABLES	viii - ix
	LIST OF GRAPHS	x - xii
CHAPTER 1		
	INTRODUCTION	1 - 3
	1.1 Overview	1
	1.2 Background	2
	1.3 Objective of the Thesis	3
	1.4 Organization of Thesis	3
CHAPTER 2		
	LITERATURE SURVEY	4 - 9
CHAPTER 3		
	SPEECH AND SPEECH PROCESSING	10 - 27
	3.1 Introduction	10
	3.2 Theory of speech production and phonetics	10
	3.2.1 Speech Production	11
	3.2.2 Phonetics	13
	3.2.3 English Anticulatory Phonetics	14
	3.3 Analysis of speech signals	15
	3.3.1 Test Preprocessing	16
	3.3.2 Pronunciation	17
	3.3.3 Prosody	18
	3.4 Speech Synthesized	19

1 Digital Filter Bank	20
2 The Fourier Transform Filter Bank	23
3 Linear Prediction Coefficients	25
 CHAPTER 4	
MATERIAL AND METHODOLOGY	28 - 41
4.1 Signal Acquisition	28
4.2 Spectrum Analysis	28
4.3 Matlab Parameters	32
4.4 SpectraPlus Parameters	36
4.5 LPC Parameters	39
 CHAPTER 5	
RESULTS AND DISCUSSION	42 - 112
5.1 Results	42
5.2 Discussion	106
5.3 Accuracy	110
 CHAPTER 6	
CONCLUSION AND FUTURE WORK	113
Conclusion	113
Future work	113
 REFERENCES	114 - 115

LIST OF FIGURES

CHAPTER 1

Figure 1.1	Process overview	2
------------	------------------	---

CHAPTER 3

Figure 3.1	Human vocal organs	11
Figure 3.2	Example of two-tube and three-tube model for the vocal tract	13
Figure 3.3	The time- and frequency-domain presentation of vowels /a/, /i/, and /u/	16
Figure 3.4	Prosodic Dependencies	18
Figure 3.5	Basic idea of the hybrid synthesis system	19
Figure 3.6	The six major spectral analysis algorithms are shown	20
Figure 3.7(a)	Bark scale	21
Figure 3.7(b)	Mel scale	21
Figure 3.7(c)	Critical Band scale	22
Figure 3.8	Spectrum shown	25

CHAPTER 4

Figure 4.1	Flow-Chat of algorithm	30
Figure 4.2	Signal of word	31
Figure 4.3	Parameters used in algorithm	32
Figure 4.4	Parameters shown	35
Figure 4.5	Spectrum	36
Figure 4.6	LPC spectrum(Red line) and Original signal spectrum(Blue line)	40
Figure 4.7	Time-frequency domain spectrum	41

CHAPTER 5

Figure 5.1	Spectrum of 'ARMY' word	43
Figure 5.2	Spectrum of 'POLICE' word	46
Figure 5.3	Spectrum of 'BOOM' word	48
Figure 5.4	Spectrum of 'BOMB' word	51
Figure 5.5	Spectrum of 'BUS' word	53
Figure 5.6	Spectrum of 'BUT' word	56

Figure 5.7	Spectrum of 'PUT' word	58
Figure 5.8	Spectrum of 'FILL' word	61
Figure 5.9	Spectrum of 'KILL' word	63
Figure 5.10	Spectrum of 'SINGH' word	66
Figure 5.11	Spectrum of 'KING' word	68
Figure 5.12	Spectrum of 'KITE' word	71
Figure 5.13	Spectrum of 'NIGHT' word	73
Figure 5.14	Spectrum of 'PUNJAB' word	76
Figure 5.15	Spectrum of 'INDIA' word	78
Figure 5.16	Spectrum of 'PAKSITAN' word	81
Figure 5.17	Spectrum of 'MINISTER' word	83
Figure 5.18	Spectrum of 'MURDER' word	86
Figure 5.19	Spectrum of 'ENGLAND' word	88
Figure 5.20	Spectrum of 'CANADA' word	91
Figure 5.21	Spectrum of 'AMERICA' word	93
Figure 5.22	Spectrum of 'AUSTRALIA' word	96
Figure 5.23	Spectrum of 'JAPAN' word	98
Figure 5.24	Spectrum of 'CHINA' word	101
Figure 5.25	Spectrum of 'RDX' word	103

LIST OF TABLES

CHAPTER 5

Table 5.1(b)	Parameters of word ‘ARMY’ by LPC Spectrum Analysis	45
Table 5.2(a)	Parameters of word ‘POLICE’ by Spectrum Analysis	46
Table 5.1(a)	Parameters of word ‘ARMY’ by Spectrum Analysis	43
Table 5.2(b)	Parameters of word ‘POLICE’ by LPC Spectrum Analysis	47
Table 5.3(a)	Parameters of word ‘BOOM’ by Spectrum Analysis	49
Table 5.3(b)	Parameters of word ‘BOOM’ by LPC Spectrum Analysis	50
Table 5.4(a)	Parameters of word ‘BOMB’ by Spectrum Analysis	51
Table 5.4(b)	Parameters of word ‘BOMB’ by LPC Spectrum Analysis	52
Table 5.5(a)	Parameters of word ‘BUS’ by Spectrum Analysis	54
Table 5.5(b)	Parameters of word ‘BUS’ by LPC Spectrum Analysis	55
Table 5.6(a)	Parameters of word ‘BUT’ by Spectrum Analysis	56
Table 5.6(b)	Parameters of word ‘BUT’ by LPC Spectrum Analysis	57
Table 5.7(a)	Parameters of word ‘PUT’ by Spectrum Analysis	59
Table 5.7(b)	Parameters of word ‘PUT’ by LPC Spectrum Analysis	60
Table 5.8(a)	Parameters of word ‘FILL’ by Spectrum Analysis	61
Table 5.8(b)	Parameters of word ‘FILL’ by LPC Spectrum Analysis	62
Table 5.9(a)	Parameters of word ‘KILL’ by Spectrum Analysis	64
Table 5.9(b)	Parameters of word ‘KILL’ by LPC Spectrum Analysis	65
Table 5.10(a)	Parameters of word ‘SINGH’ by Spectrum Analysis	66
Table 5.10(b)	Parameters of word ‘SINGH’ by LPC Spectrum Analysis	67
Table 5.11(a)	Parameters of word ‘KING’ by Spectrum Analysis	69
Table 5.11(b)	Parameters of word ‘KING’ by LPC Spectrum Analysis	70
Table 5.12(a)	Parameters of word ‘KITE’ by Spectrum Analysis	71
Table 5.12(b)	Parameters of word ‘KITE’ by LPC Spectrum Analysis	72
Table 5.13(a)	Parameters of word ‘NIGHT’ by Spectrum Analysis	74
Table 5.13(b)	Parameters of word ‘NIGHT’ by LPC Spectrum Analysis	75
Table 5.14(a)	Parameters of word ‘PUNJAB’ by Spectrum Analysis	76
Table 5.14(b)	Parameters of word ‘PUNJAB’ by LPC Spectrum Analysis	77
Table 5.15(a)	Parameters of word ‘INDIA’ by Spectrum Analysis	79
Table 5.15(b)	Parameters of word ‘INDIA’ by LPC Spectrum Analysis	80

Table 5.16(a)	Parameters of word ‘PAKSITAN’ by Spectrum Analysis	81
Table 5.16(b)	Parameters of word ‘PAKSITAN’ by LPC Spectrum Analysis	82
Table 5.17(a)	Parameters of word ‘MINISTER’ by Spectrum Analysis	84
Table 5.17(b)	Parameters of word ‘MINISTER’ by LPC Spectrum Analysis	85
Table 5.18(a)	Parameters of word ‘MURDER’ by Spectrum Analysis	86
Table 5.18(b)	Parameters of word ‘MURDER’ by LPC Spectrum Analysis	87
Table 5.19(a)	Parameters of word ‘ENGLAND’ by Spectrum Analysis	89
Table 5.19(b)	Parameters of word ‘ENGLAND’ by LPC Spectrum Analysis	90
Table 5.20(a)	Parameters of word ‘CANADA’ by Spectrum Analysis	91
Table 5.20(b)	Parameters of word ‘CANADA’ by LPC Spectrum Analysis	92
Table 5.21(a)	Parameters of word ‘AMERICA’ by Spectrum Analysis	94
Table 5.21(b)	Parameters of word ‘AMERICA’ by LPC Spectrum Analysis	95
Table 5.22(a)	Parameters of word ‘AUSTRALIA’ by Spectrum Analysis	96
Table 5.22(b)	Parameters of word ‘AUSTRALIA’ by LPC Spectrum Analysis	97
Table 5.23(a)	Parameters of word ‘JAPAN’ by Spectrum Analysis	99
Table 5.23(b)	Parameters of word ‘JAPAN’ by LPC Spectrum Analysis	100
Table 5.24(a)	Parameters of word ‘CHINA’ by Spectrum Analysis	101
Table 5.24(b)	Parameters of word ‘CHINA’ by LPC Spectrum Analysis	102
Table 5.25(a)	Parameters of word ‘RDX’ by Spectrum Analysis	104
Table 5.25(b)	Parameters of word ‘RDX’ by LPC Spectrum Analysis	105
Table 5.26	Parameters show their LDA for words ‘FILL-KILL’	107
Table 5.27	Parameters Show their LDA for words ‘BOOM-BOMB’	108
Table 5.28	Parameters show their LDA for words ‘BUT-PUT’	110
Table 5.29	Identified words	111

LIST OF GRAPHS

CHAPTER 5

Graph 5.1(a)	Parameters show their range of values from table 5.1(a)	44
Graph 5.1(b)	Insignificant Parameters from table 5.1(a)	44
Graph 5.1(c)	Significant Parameters from table 5.1(a)	44
Graph 5.1(d)	Significant Parameters from table 5.1(b)	45
Graph 5.2(a)	Insignificant Parameters from table 5.2(a)	47
Graph 5.2(b)	Significant Parameters from table 5.2(a)	47
Graph 5.2(c)	Significant Parameters from table 5.2(b)	48
Graph 5.3(a)	Insignificant Parameters from table 5.3(a)	49
Graph 5.3(b)	Significant Parameters from table 5.3(a)	50
Graph 5.3(c)	Significant Parameters from table 5.3(b)	50
Graph 5.4(a)	Insignificant Parameters from table 5.4(a)	52
Graph 5.4(b)	Significant Parameters from table 5.4(a)	52
Graph 5.4(c)	Significant Parameters from table 5.4(b)	53
Graph 5.5(a)	Insignificant Parameters from table 5.5(a)	54
Graph 5.5(b)	Significant Parameters from table 5.5(a)	55
Graph 5.5(c)	Significant Parameters from table 5.5(b)	55
Graph 5.6(a)	Insignificant Parameters from table 5.6(a)	57
Graph 5.6(b)	Significant Parameters from table 5.6(a)	57
Graph 5.6(c)	Significant Parameters from table 5.6(b)	58
Graph 5.7(a)	Insignificant Parameters from table 5.7(a)	59
Graph 5.7(b)	Significant Parameters from table 5.7(a)	60
Graph 5.7(c)	Significant Parameters from table 5.7(b)	60
Graph 5.8(a)	Insignificant Parameters from table 5.8(a)	62
Graph 5.8(b)	Significant Parameters from table 5.8(a)	62
Graph 5.8(c)	Significant Parameters from table 5.8(b)	63
Graph 5.9(a)	Insignificant Parameters from table 5.9(a)	64
Graph 5.9(b)	Significant Parameters from table 5.9(a)	65
Graph 5.9(c)	Significant Parameters from table 5.9(b)	65
Graph 5.10(a)	Insignificant Parameters from table 5.10(a)	67
Graph 5.11(b)	Significant Parameters from table 5.10(a)	67

Graph 5.11(c)	Significant Parameters from table 5.10(b)	68
Graph 5.12(a)	Insignificant Parameters from table 5.12(a)	72
Graph 5.12(b)	Significant Parameters from table 5.12(a)	72
Graph 5.12(c)	Significant Parameters from table 5.12(b)	73
Graph 5.13(a)	Insignificant Parameters from table 5.13(a)	74
Graph 5.13(b)	Significant Parameters from table 5.13(a)	75
Graph 5.13(c)	Significant Parameters from table 5.13(b)	75
Graph 5.14(a)	Insignificant Parameters from table 5.14(a)	77
Graph 5.14(b)	Significant Parameters from table 5.14(a)	77
Graph 5.14(c)	Significant Parameters from table 5.14(b)	78
Graph 5.15(a)	Insignificant Parameters from table 5.15(a)	79
Graph 5.15(b)	Significant Parameters from table 5.15(a)	80
Graph 5.15(c)	Significant Parameters from table 5.15(b)	80
Graph 5.16(a)	Insignificant Parameters from table 5.16(a)	82
Graph 5.16(b)	Significant Parameters from table 5.16(a)	82
Graph 5.16(c)	Significant Parameters from table 5.16(b)	83
Graph 5.17(a)	Insignificant Parameters from table 5.17(a)	84
Graph 5.17(b)	Significant Parameters from table 5.17(a)	85
Graph 5.17(c)	Significant Parameters from table 5.17(b)	85
Graph 5.18(a)	Insignificant Parameters from table 5.18(a)	87
Graph 5.18(b)	Significant Parameters from table 5.18(a)	87
Graph 5.18(c)	Significant Parameters from table 5.18(b)	88
Graph 5.19(a)	Insignificant Parameters from table 5.19(a)	89
Graph 5.19(b)	Significant Parameters from table 5.19(a)	90
Graph 5.19(c)	Significant Parameters from table 5.19(b)	90
Graph 5.20(a)	Insignificant Parameters from table 5.20(a)	92
Graph 5.20(b)	Significant Parameters from table 5.20(a)	92
Graph 5.20(c)	Significant Parameters from table 5.20(b)	93
Graph 5.21(a)	Insignificant Parameters from table 5.21(a)	94
Graph 5.21(b)	Significant Parameters from table 5.21(a)	95
Graph 5.21(c)	Significant Parameters from table 5.21(b)	95
Graph 5.22(a)	Insignificant Parameters from table 5.22(a)	97
Graph 5.22(b)	Significant Parameters from table 5.22(a)	97

Graph 5.22(c)	Significant Parameters from table 5.22(b)	98
Graph 5.23(a)	Insignificant Parameters from table 5.23(a)	99
Graph 5.23(b)	Significant Parameters from table 5.23(a)	100
Graph 5.23(c)	Significant Parameters from table 5.23(b)	100
Graph 5.24(a)	Insignificant Parameters from table 5.24(a)	102
Graph 5.24(b)	Significant Parameters from table 5.24(a)	102
Graph 5.24(c)	Significant Parameters from table 5.24(b)	103
Graph 5.25(a)	Insignificant Parameters from table 5.25(a)	104
Graph 5.25(b)	Significant Parameters from table 5.25(a)	105
Graph 5.25(c)	Significant Parameters from table 5.25(b)	105
Graph 5.26	Differentiate the word by some parameters	106
Graph 5.27(a)	Differentiate the word by some parameters	107
Graph 5.27(b)	Differentiate the word by P.frq parameter	108
Graph 5.27(c)	Differentiate the word by Slop parameter	108
Graph 5.28(a)	Differentiate the word by P.Frq parameter	109
Graph 5.28(b)	Differentiate the word by THD+N parameter	109
Graph 5.28(c)	Differentiate the word by some parameters	109
Graph 5.28(d)	Differentiate the word by T.Power parameter	110

CHAPTER 1

INTRODUCTION

1.1 Overview

Speech recognition is a multileveled pattern recognition task, in which acoustical signals are examined and structured into a hierarchy of sub word units (e.g., phonemes), words, phrases, and sentences. Each level provides additional temporal constraints, e.g., known word pronunciations or legal word sequences, which can compensate for errors or uncertainties at lower levels. This thesis demonstrates method to recognize the words from a speech signal.

To accomplish the above objective speech signal spectrum is used. First of all some words are recorded by recorder and converted into the wave file format. To recognize the word, spectrum based parameters are found. When analysis of the speech signal spectrum is done, near about twenty four parameters are obtained. Some of the parameters are Mean, Median, Standard Deviation (STD), Root Mean Square(RMS), Maximum Peak, Minimum Peak, Slope of maximum peak, Width of maximum peak, Signal to Noise Ratio(SNR), Peak Frequency, Peak Amplitude, Total Power, Total Harmonic Distortion(THD), THD + Noise, Inter Modulation Distortion (IMD). Samples of some words are analyzed by statistical methods which are given some range of values. Every word parameter has its own value and lies in between some bounded or range of values.

Finding the most informative parameters of speech signal is the key factor to improve the recognition of word. To find these parameters some statistical methods and second technique Linear Predicted Coding (LPC) coefficients are used. From this method new signal spectrum is obtained and it is different from the original speech signal spectrum. So by this new spectrum analysis another new set of parameters are found which helps us to recognize the word. So in this work, twenty five words each having five samples are collected and analysis of their signal spectrum is done to find various parameters. From these parameter values the word is recognized.

Speech should be initially transformed and compressed, in order to simplify subsequent processing. Many signal analysis techniques are available which can extract useful features and

compress the data by a factor of ten without losing any important information. Among the most popular methods are FFT and LPC. Word signals are converted into the spectrum signal using FFT but FFT requires only complex values. Similarly by using LPC spectrum program we get different spectrum from the original spectrum and then analysis on their spectrum is done to find other parameters. The following steps of all this work as shown in figure 1.1. There are various applications of word recognition, like telephone talking, on-line and off-line communications etc.

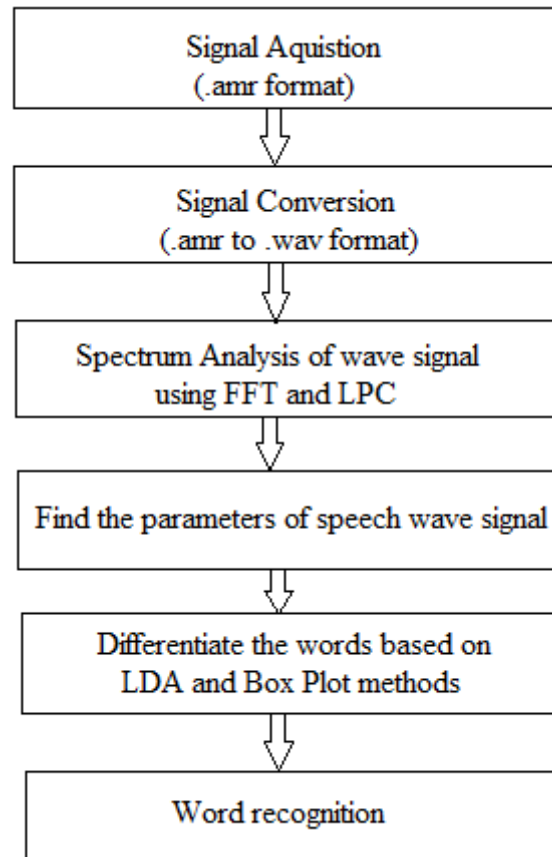


Figure 1.1: Process overview

1.2 Background

Speech is the most natural form of human communication. Speech processing has been one of the most exciting areas of signal processing. Speech (word) recognition technology has made it possible for computer to follow human voice commands and even understand human languages. To recognize the word there are various method which are already used. Mel-frequency cepstrum coefficients (MFCC) parameters are used for recognition of the word. LPC

parameter is also used in recognition of the word. This work is on LPC spectrum analysis and various parameters are obtained.

- Hidden Filter Model are used for finding the parameters such as MFCC and LPC. Also used for parameter selection and sensitivity to power normalization.
- Gaussian Mixture Model are used for MFCC.
- GUI-Based tool are used for speech signal processing which are finding the parameters.
- Artificial Neural Networking used for finding the parameters of word.

1.3 Objective of the Thesis

Word recognition is the main objective of this thesis. To accomplish this task, spectrum of the speech signal is analyzed by FFT and LPC techniques and various parameters are found. By getting the parameter values correct word recognition is done.

1.4 Organization of Thesis

- Chapter 1 “**INTRODUCTION**”, it includes the introduction about the overview and objective of the work done in this thesis.
- Chapter 2 “**LITERATURE SURVEY**”, it provides the overview of the work done in this area.
- Chapter 3 “**SPEECH AND SPEECH PROCEEING**”, it gives the basic idea of speech and which are useful in speech processing and analysis on spectrum.
- Chapter 4 “**MATERIALS AND METHODLOGY**”, how to take the material and explains the analysis of the speech signal spectrum by various methods and find parameters.
- Chapter 6 “**RESULTS AND DISSCUSSION**”, get the parameters by analysis the speech signal spectrum.
- Chapter 7 “**CONCLUSION AND FUTURE WORK**”, conclude thesis with future scope.

CHAPTER 2

LITERATURE SURVEY

J. D. Tardelli , C. M. Walter ; describes the implementation of a matrix array processor oriented technique that can transform time sampled speech waveform data, by way of the frequency domain into a pseudo-canonical coordinate domain whose components can be ordered. The components are ordered in such a manner so as to commit the smallest signal reconstruction error, relative to a given non-Euclidean quadratic error criterion. The error metric can be specified, in either time or frequency domain, in the form of a complex valued hermitian matrix, thus allowing a very less number of degrees of freedom that can be used to in statistical, dynamical and psychoacoustic attributes of the speech data into the analysis and reconstruction process. For the case of a simple Euclidean error metric, the pseudo-canonical coordinate process is shown to reduce to the well known principal component, signal representation technique. [1]

Takao Suzuki, Yasuo Shoji ; from the viewpoint of speech communication services for Asynchronous Transfer Mode (ATM) network and in order to introduce the necessary conditions for speech processing over an ATM network, a new speech processing scheme applied at the end of the ATM network. For this speech processing, speech signals are processed basically by two techniques; these are silence deletion for speech compression and low-bit speech coding for 32-kbi vs Adaptive Differential PCM (ADPCM). In order to reduce speech quality degradation caused by lost ATM cells in network congestion conditions, a new cell reconstruction algorithm using waveform substitution for ADPCM-coded speech based on the pitch estimation method. In addition, to maintain good speech quality some new algorithms for speech processing is used. Therefore through subjective evaluation tests that the proposed speech processing scheme for the ATM network could provide good speech quality up to a cell loss rate of about 3% for public communications. [2]

Hamid Sheikhzadeh, Li Deng ; describes a novel approach to speech recognition by directly modeling the statistical characteristics of the speech waveforms. This approach allows us to remove the need for using speech preprocessors, which conventionally serve a role of converting speech waveforms into frame-based speech data subject to a subsequent modeling process. Central to our method is the representation of the speech waveforms as the output of a

time-varying filter excited by a Gaussian source time-varying in its power. In order to formulate a speech recognition algorithm based on this representation, the time variation in the characteristics of the filter and of the excitation source is described in a compact and parametric form of the Markov chain. Detail the comparative roles played by the filter modeling and by the source modeling in speech recognition performance. Based on their results of the analysis, here evaluate a normalization procedure intended to remove the sensitivity of speech recognition accuracy to often uncontrollable speech power variations. Effectiveness of the proposed speech-waveform modeling approach is demonstrated in a speaker-dependent, discrete-utterance speech recognition task involving 18 highly confusable stop consonant-vowel syllables. The high accuracy obtained shows promising potentials of the proposed time-domain waveform modelling technique for speech recognition. [3]

Mustafa Nazmi Kaynak, Qi Zhi ; audio-visual speech recognition is a novel extension of acoustic speech recognition and has received a lot of attention in the last few decades. The main motivation behind bimodal speech recognition is the bimodal characteristics of speech perception and production systems of human beings. The effect of the modeling parameters of hidden markov models (HMM) on the recognition accuracy of the bimodal speech recognizer is analyzed, a comparative analysis of the different HMMs that can be used in bimodal speech recognition is presented, and finally a novel model, which has been experimentally verified to perform better than other is proposed. Also the geometric visual features are compared and analyzed for their importance in bimodal speech recognition. One of the unique characteristics of our bimodal speech recognition system is the novel fusion strategy of the acoustic and the visual features, which takes into account the different sampling rates of these two signals. Compared to acoustic only, the audio-visual speech recognition scheme has much more improved recognition accuracy, especially in presence of noise. [5]

Jose Leonardo Plaza-Aguilar, David Baez-Lopez ; to show how MATLAB can be used to realize Voice Recognition; for people with Cerebral Paralysis. The first part is about what is the Cerebral Paralysis, and its consequences in the way of talk of the people with this illness. After this, we will mention the mathematic background needed to realize this project besides of explaining a first stage of the system, on this stage we could recognize some

phonemes. Next, we will describe the functioning of the second stage of the project. On this part, we could recognize words. [6]

Vikas C. Raykar , S. R. Mahadeva Prasanna ; to studies for localization of a moving speaker using information about the excitation source of speech production. The first step in localization is the estimation of time-delay from speech collected by a pair of microphones. Methods for time-delay estimation generally use spectral features that correspond mostly to the shape of vocal tract during speech production. Spectral features are affected by degradations due to noise and reverberation. This is proposes a method for localizing a speaker using features that arise from the excitation source during speech production. Experiments were conducted by simulating different noise and reverberation conditions to compare the performance of the time-delay estimation and source localization using the proposed method with the results obtained using the spectrum-based generalized cross correlation (GCC) methods. The bias, variance and the root mean square error (RMSE) of the proposed method is consistently equal or less than the GCC methods. The location of a moving speaker estimated using the time-delays obtained by the proposed method are closer to the actual values, than those obtained by the GCC method. [7]

Marcus Holmberg, David Gelbart ; the Mel-frequency Cepstral Coefficient (MFCC) or perceptual linear prediction (PLP) feature extraction typically used for automatic speech recognition (ASR) employ several principles which have known counterparts in the cochlea and auditory nerve: frequency decomposition, Mel-frequency or bark-warping of the frequency axis, and compression of amplitudes. It seems natural to ask if one can profitably employ a counterpart of the next physiological processing step, synaptic adaptation. Therefore incorporated a simplified model of short-term adaptation into MFCC feature extraction. Evaluated the resulting ASR performance on the AURORA 2 and AURORA 3 tasks, in comparison to ordinary MFCCs, MFCCs processed by RASTA, and MFCCs processed by cepstral mean subtraction (CMS), and both in comparison to and in combination with Wiener filtering. The results suggest that it offers a simple, causal robustness strategy which is competitive with RASTA, CMS, and Wiener filtering and performs well in combination with Wiener filtering. Compared to the structurally related RASTA, this adaptation model provides superior performance on AURORA 2 and if Wiener filtering is used prior to both approaches, on AURORA 3 as well. [8]

Marco Grimaldi, Fred Cummins ; an experimental evaluation of different features for use in speaker identification. The features are tested using speech data provided by the CHAINS corpus, in a closed-set speaker identification task. The main objective is to present a novel parameterization of speech that is based on the AM–FM representation of the speech signal and to assess the utility of these features in the context of speaker identification. In order to explore the extent to which different instantaneous frequencies due to the presence of formants and harmonics in the speech signal may predict a speaker's identity, this work evaluates three different decompositions of the speech signal within the same AM–FM framework: a first setup has been used previously for formant tracking, a second setup is designed to enhance familiar resonances below 4000 Hz, and a third setup is designed to approximate the bandwidth scaling of the filters conventionally used in the extraction of Mel-frequency cepstral coefficients (MFCCs). From each of the proposed setups, parameters are extracted and used in a closed-set text-independent speaker identification task. The performance of the new fractural representation is compared with results obtained adopting MFCC and RASTA-PLP features in the context of a generic Gaussian mixture model (GMM) classification system. In evaluating the novel features, we look selectively at information for speaker identification contained in the frequency range 0–4000 Hz and 4000–8000 Hz, as the instantaneous frequencies revealed by the AM–FM approach suggest the presence of structures not well known conventional spectrographic analyses. [9]

Ibrahim N. Abu-Isbeih , Khaled Daqrouq , Wael Al-Sawalmeh ; one of the most important signal processing method in Digital Signal Processing discipline is Speaker Identification Method (SIM). Because of the difficult nature of speech signals and their fast variation with time, the wavelet transform is used to reduce the complexity of such signals. In this two identification methods are presented based on Continuous Wavelet Transform CWT. The first method is the male and female identification method and the second is the SIM. The speaker identification wavelet transform based method to be very important in digital security systems applications. [10]

Dr. E. Chandra, Mrs. C. Sunitha ; discuss a speech-and-speaker (SAS) identification system. The speech signal is treated graphically in order to extract the essential image features as a basic step in successful data mining applications in the biometric techniques. The main object here is the human-voice signal. The identifying and classifying methods are performed with

Burg's estimation model and the algorithm of Toilets matrix minimal Eigen values is used as the main tools for signal-image description and feature extraction. The extracted feature-carrying image comprises the elements of Toilets matrices to consecutively compute their minimal Eigen values and introduce a set of feature vectors within a class of voices. At the stage of classification, both conventional and neural-network-based methods are used. This helps in speech recognition and speaker authentication. [11]

Anup Kumar Paul, Dipankar Das, Md. Mustafa Kamal ; they describe the Bangla speech recognition system. Bangla speech recognition system is divided mainly into two major parts. The first part is speech signal processing and the second part is speech pattern recognition technique. The speech processing stage consists of speech starting and end point detection, windowing, filtering, calculating the Linear Predictive Coding (LPC) and Cepstral Coefficients and finally constructing the codebook by vector quantization. The second part consists of pattern recognition system using Artificial Neural Network (ANN). The recorded speech signal is passed through the speech starting and end-point detection algorithm to detect the presence of the speech signal and remove the silence and pauses portions of the signals. The resulting signal is then filtered for the removal of unwanted background noise from the speech signals. The filtered signal is then windowed ensuring half frame overlap. After windowing, the speech signal is then subjected to calculate the LPC coefficient and Cepstral coefficient. The feature extractor uses a standard LPC Cepstrum coder, which converts the incoming speech signal into LPC Cepstrum feature space. The Self Organizing Map(SOM) Neural Network makes each variable length LPC trajectory of an isolated word into a fixed length LPC trajectory and thereby making the fixed length feature vector, to be fed into to the recognizer. The structures of the neural network is designed with Multi Layer Perceptron approach and tested with 3, 4, 5 hidden layers using the Transfer functions of Tan Sigmoid for the Bangla speech recognition system. Comparison among different structures of Neural Networks conducted here for a better understanding of the problem and its possible solutions. [12]

Zufeng Weng, Donghui Guo ; a new algorithm of feature parameter extraction is proposed for application in speaker recognition system, which combines the traditional MFCC and the dynamic MFCC as a new series of coefficients. According to the statistics analysis of the different contribution by the dynamic MFCC and traditional MFCC, these coefficients are

weighted as front-end parameters of the GMM, which would decrease the dimension of the mixed weighted GMM and reduce the computation complexity. The experiments based on the speech database were implemented in MATLAB environment, and the results showed the speaker recognition system with the Weighted Dynamic MFCC could obtain better performance with high recognition rate and low computational complexity. [13]

Shweta Ghai, Rohit Sinha ; the degradation in children's speech recognition performance under mismatched condition i.e., on the adults' speech trained models is a well known problem. Apart from several other factors, this degradation is also contributed by the large difference in the pitch values of the adults' and the children's speech. MFCC is the most commonly used feature in automatic speech recognition but it has been reported to be affected by the pitch variations across speech signals. Recently, perceptual-MVDR (PMVDR) feature has been reported as a better alternative to MFCC under noisy conditions. It is also attributed to possess better spectral modeling ability for high pitch signals. In this, analyze the robustness of PMVDR to pitch variations across speech signals in comparison to MFCC for the children's speech recognition under mismatched condition. Finds PMVDR to be more pitch robust than MFCC using the default parameters. However, on suitable adaptation of the parameters for the children's speech recognition under mismatched condition, both PMVDR and MFCC give significantly improved comparable performances for children's speech as well as exhibit similar robustness to pitch variations. [14]

Takahiro Murakami, Yoshihisa Ishida ; a new method for fundamental frequency estimation from the noisy spectrum of a speech signal is introduced. The fundamental frequency is one of the most essential characteristics for speech recognition, speech coding and so on. The proposed method uses the MUSIC algorithm, which is an Eigen-based subspace decomposition method. [15]

3.1 Introduction

Knowledge extraction by just listening to sounds is a distinctive property and has become an important milestone in the evolution of species. Most of the animals are not only equipped with the means to extract information from the rich acoustical content of the environment and act accordingly, but they have the ability to produce sounds to interact with the environment as well. Humans have gone one step further; they have fairly advanced mechanisms that enable interaction within the species by very abstract rules of communication using voice – the language.

Speech is the primary means of communication between people. Speech synthesis, automatic generation of speech waveforms, has been under development for several decades. Recent progress in speech synthesis has produced synthesizers with very high intelligibility but the sound quality and naturalness still remain a major problem. However, the quality of present products has reached an adequate level for several applications, such as multimedia and telecommunications. With some audiovisual information or facial animation (talking head) it is possible to increase speech intelligibility considerably. Preparing for research in voice synthesis requires a fairly diverse range of coursework and understanding. Many biological and physiological functions must be understood for use as a background. In order to synthesize human speech, an understanding must be developed of how humans are capable of such a task. By dissecting speech and its process into measurable and analytical qualities, instruments and techniques can be created to reproduce the aspects of speech.

3.2 Theory of Speech Production and Phonetics

Speech processing and language technology contains lots of special concepts and terminology. To understand how different speech synthesis and analysis methods work one must have some knowledge of speech production, articulator phonetics, and some other related terminology. The basic theory of these topics will be discussed briefly in this topic.

3.2.1 Speech Production

Human speech is produced by vocal organs presented in Figure 3.1. The main energy source is the lungs with the diaphragm. [22] When speaking, the air flow is forced through the glottis between the vocal cords and the larynx to the three main cavities of the vocal tract, the pharynx and the oral and nasal cavities. From the oral and nasal cavities the air flow exits through the nose and mouth, respectively. The V-shaped opening between the vocal cords, called the glottis, is the most important sound source in the vocal system.

The vocal cords may act in several different ways during speech. The most important function is to modulate the air flow by rapidly opening and closing, causing buzzing sound from which vowels and voiced consonants are produced. The fundamental frequency of vibration depends on the mass and tension and is about 110 Hz, 200 Hz, and 300 Hz with men, women, and children, respectively. With stop consonants the vocal cords may act suddenly from a completely closed position, in which they cut the air flow completely, to totally open position producing a light cough or a glottal stop. On the other hand, with unvoiced consonants, such as /s/ or /f/, they may be completely open. An intermediate position may also occur with for example phonemes like /h/.

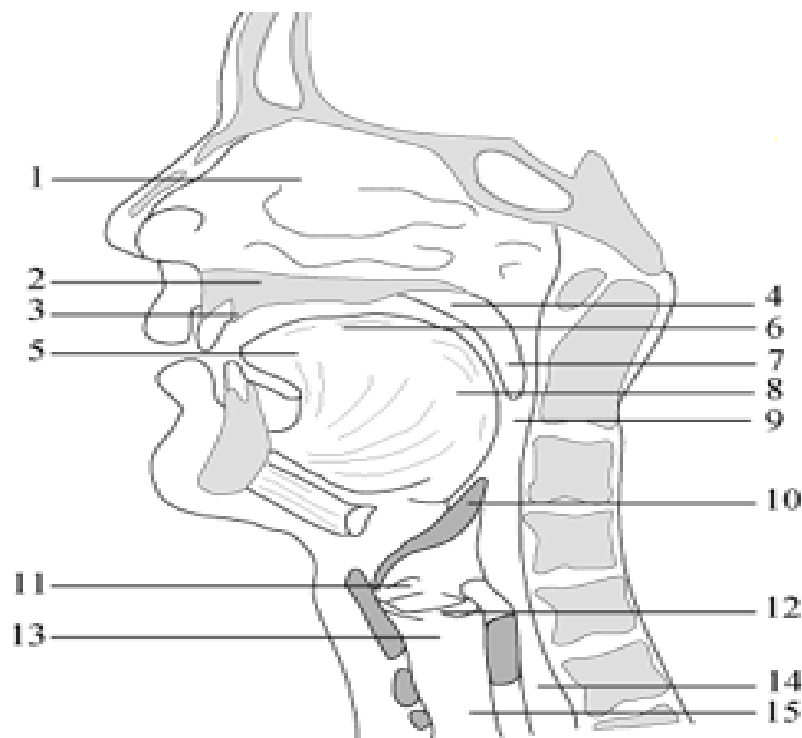


Figure 3.1: Human vocal organs

(1) Nasal cavity, (2) Hard palate, (3) Alveolar ridge, (4) Soft palate (Velum), (5) Tip of the tongue (Apex), (6) Dorsum, (7) Uvula, (8) Radix, (9) Pharynx, (10) Epiglottis, (11) False vocal cords, (12) Vocal cords, (13) Larynx, (14) Esophagus, and (15) Trachea.

The pharynx connects the larynx to the oral cavity. It has almost fixed dimensions, but its length may be changed slightly by raising or lowering the larynx at one end and the soft palate at the other end. The soft palate also isolates or connects the route from the nasal cavity to the pharynx. At the bottom of the pharynx are the epiglottis and false vocal cords to prevent food reaching the larynx and to isolate the esophagus acoustically from the vocal tract. The epiglottis, the false vocal cords and the vocal cords are closed during swallowing and open during normal breathing. The oral cavity is one of the most important parts of the vocal tract. Its size, shape and acoustics can be varied by the movements of the palate, the tongue, the lips, the cheeks and the teeth. Especially the tongue is very flexible, the tip and the edges can be moved independently and the entire tongue can move forward, backward, up and down. The lips control the size and shape of the mouth opening through which speech sound is radiated.

Unlike the oral cavity, the nasal cavity has fixed dimensions and shape. Its length is about 12 cm and volume 60 cm³. The air stream to the nasal cavity is controlled by the soft palate. From technical point of view, the vocal system may be considered as a single acoustic tube between the glottis and mouth. Glottal excited vocal tract may be then approximated as a straight pipe closed at the vocal cords where the acoustical impedance $Z_g = \infty$ and open at the mouth ($Z_m = 0$). In this case the volume-velocity transfer function of vocal tract is:

$$V(\omega) = Z_m/Z_g = U_m/U_g = 1/\cos(\omega l/c) \quad (1)$$

Where l is the length of the tube, ω is radian frequency and c is sound velocity.

Vowels can be approximated with a two-tube model presented on the left in Figure 3.2. For example, with vowel /a/ the narrower tube represents the pharynx opening into wider tube representing the oral cavity. If assumed that both tubes have an equal length of 8.5cm, formants occur at twice the frequencies noted earlier for a single tube. Due to acoustic coupling, formants do not approach each other by less than 200 Hz so formants F1 and F2 for /a/ are not both at 1000 Hz, but rather 900 Hz and 1100 Hz, respectively.

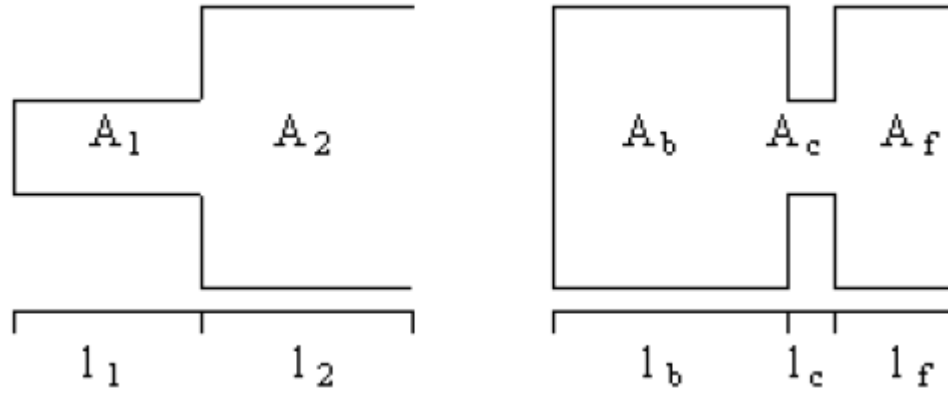


Figure 3.2: Examples of two-tube and three-tube models for the vocal tract

Consonants can be approximated similarly with a three-tube model shown on the right in Figure 3.2, where the narrow middle tube models the vocal tract constriction. The back and middle tubes are half-wavelength resonators and the front tube is a quarter wavelength resonators with resonances:

$$C_i/2l_b, C_i/2l_c, C(2i-1)/4l_f, \quad i = 1, 2, 3, \dots \quad (2)$$

Where l_b , l_c and l_f are the length of the back, center and front tube respectively. The excitation signal may be modeled with a two-mass model of the vocal cords which consists of two masses coupled with a spring and connected to the larynx by strings and damper. Several other methods and systems have been developed to model the human speech production system to produce synthetic speech.

3.2.2 Phonetics

In most languages the written text does not correspond to its pronunciation so that in order to describe correct pronunciation some kind of symbolic presentation is needed. Every language has a different phonetic alphabet and a different set of possible phonemes and their combinations. The number of phonetic symbols is between 20 and 60 in each language. A set of phonemes can be defined as the minimum number of symbols needed to describe every possible word in a language. In English there are about 40 phonemes. Due to complexity and different kind of definitions, the number of phonemes in English and most of the other languages cannot be defined exactly.

Phonemes are abstract units and their pronunciation depends on contextual effects, speaker's characteristics, and emotions. During continuous speech, the articulatory movements depend on the preceding and the following phonemes. The articulators are in different position depending on the preceding one and they are preparing to the following phoneme in advance. This causes some variations on how the individual phoneme is pronounced. These variations are called allophones which are the subset of phonemes and the effect is known as coarticulation. For example, a word *lice* contains a light /l/ and *small* contains a dark /l/. These l's are the same phoneme but different allophones and have different vocal tract configurations. Another reason why the phonetic representation is not perfect is that the speech signal is always continuous and phonetic notation is always discrete.

Different emotions and speaker characteristics are also impossible to describe with phonemes so the unit called phone is usually defined as an acoustic realization of a phoneme. The phonetic alphabet is usually divided in two main categories, vowels and consonants. Vowels are always voiced sounds and they are produced with the vocal cords in vibration, while consonants may be either voiced or unvoiced. Vowels have considerably higher amplitude than consonants and they are also more stable and easier to analyze and describe acoustically. Because consonants involve very rapid changes they are more difficult to synthesize properly. Some efforts to construct language-independent phonemic alphabets were made during last decades. One of the best known is perhaps IPA (International Phonetic Alphabet) which consists of a huge set of symbols for phonemes, super segmental, tones/word accent contours, and diacritics. For example, there are over twenty symbols for only fricative consonants.

3.2.3 English Articulatory Phonetics

The number of phonetic symbols used in English varies by different kind of definitions. Usually there are about ten to fifteen vowels and about twenty to twenty-five consonants. English vowels may be classified by the manner or place of articulation (front-back) and by the shape of the mouth (open - close). English consonants may be classified by the manner of articulation as plosives, fricatives, nasals, liquids, and semivowels. Plosives are known also as stop consonants. Liquids and semivowels are also defined in some publications as approximants and laterals. Further classification may be made by the place of articulation as labials (lips),

dentals (teeth), alveolar (gums), palatals (palate), velars (soft palate), glottal (glottis), and labiodentals (lips and teeth).

3.3 Analysis of Speech Signals

Continuous speech is a set of complicated audio signals which makes producing them artificially difficult. Speech signals are usually considered as voiced or unvoiced, but in some cases they are something between these two. Voiced sounds consist of fundamental frequency (F_0) and its harmonic components produced by vocal cords (vocal folds).[21] The vocal tract modifies this excitation signal causing formant (pole) and sometimes ant formant (zero) frequencies .Each formant frequency has also amplitude and bandwidth and it may be sometimes difficult to define some of these parameters correctly. The fundamental frequency and formant frequencies are probably the most important concepts in speech synthesis and also in speech processing in general. With purely unvoiced sounds, there is no fundamental frequency in excitation signal and therefore no harmonic structure either and the excitation can be considered as white noise. The airflow is forced through a vocal tract constriction which can occur in several places between glottis and mouth. Some sounds are produced with complete stoppage of airflow followed by a sudden release, producing an impulsive turbulent excitation often followed by a more protracted turbulent excitation [10]. Unvoiced sounds are also usually more silent and less steady than voiced ones.

Speech signals of the three vowels (/a/ /i/ /u/) are presented in time- and frequency domain in Figure 3.3. The fundamental frequency is about 100 Hz in all cases and the formant frequencies F_1 , F_2 , and F_3 with vowel /a/ are approximately 600 Hz, 1000 Hz, and 2500 Hz respectively. With vowel /i/ the first three formants are 200 Hz, 2300 Hz, and 3000 Hz, and with /u/ 300 Hz, 600 Hz, and 2300 Hz. The harmonic structure of the excitation is also easy to perceive from frequency domain presentation.

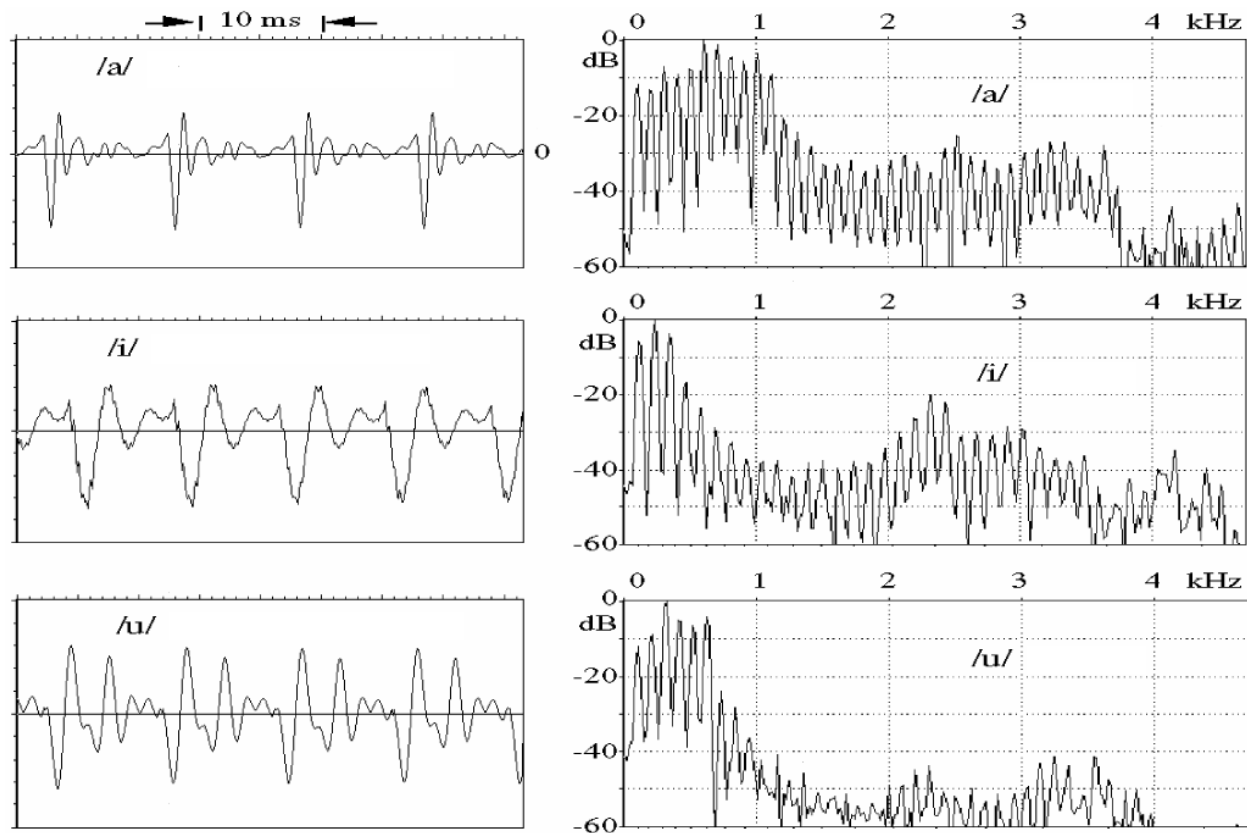


Figure 3.3: The time- and frequency-domain presentation of vowels /a/, /i/, and /u/.

It can be seen that the first three formants are inside the normal telephone channel (from 300 Hz to 3400 Hz) so the needed bandwidth for intelligible speech is not very wide. For higher quality, up to 10 kHz bandwidth may be used which leads to 20 kHz sampling frequency. Unless, the fundamental frequency is outside the telephone channel, the human hearing system is capable to reconstruct it from its harmonic components.

3.3.1 Text preprocessing

Text preprocessing is usually a very complex task and includes several language dependent problems. Digits and numerals must be expanded into full words. For example in English, numeral 243 would be expanded as two hundred and forty-three and 1750 as seventeen-fifty (if year) or one-thousand seven-hundred and fifty (if measure). Fractions and dates are also problematic. 5/16 can be expanded as five-sixteenths (if fraction) or May sixteenth (if date). Expansion ordinal numbers have been found also problematic.

The first three ordinals must be expanded differently than the others, 1st as first, 2nd as second, and 3rd as third. Same kind of contextual problems are faced with roman numerals.

Numbers may also have some special forms of expression, such as 22 as double two in telephone numbers and 1-0 as one love in sports. Abbreviations may be expanded into full words, pronounced as written or pronounced letter by letter. Some contextual problems for example kg can be either kilogram or kilograms depending on preceding number, St. can be saint or street and ft. Fort, foot or feet. In some cases, the adjacent information may be enough to find out the correct conversion, but to avoid miss conversions the best solution in some cases may be the use of letter-to-letter conversion.

3.3.2 Pronunciation

The second task is to find correct pronunciation for different contexts in the text. Some words, called homographs, cause maybe the most difficult problems in TTS systems. Homographs are spelled the same way but they differ in meaning and usually in pronunciation (e.g. fair, lives). The word lives is for example pronounced differently in sentences "Three lives were lost" and "One lives to eat". Some words e.g. lead, has different pronunciations when used as a verb or noun, and between two noun senses (He followed her lead / He covered the hull with lead). With these kinds of words some semantically information is necessary to achieve correct pronunciation.

The pronunciation of a certain word may also be different due to contextual effects. This is easy to see when comparing phrases the end and the beginning. The pronunciation depends on the initial phoneme in the following word. Compound words are also problematic. For example the characters 'th' in mother and hothouse is pronounced differently. Some sounds may also be either voiced or unvoiced in different context. For example, phoneme /s/ in word dogs is voiced, but unvoiced in word cats. Finding correct pronunciation for proper names, especially when they are borrowed from other languages, is usually one of the most difficult tasks for any TTS system. Some common names, such as Nice and Begin, are ambiguous in capitalized context, including sentence initial position, titles and single text. For example, the sentence Nice is a nice place is very problematic because the word Nice may be pronounced as /niis/ or /nais/. Some names and places have also special pronunciation, such as Leicester and Arkansas. For correct pronunciation, these kinds of words may be included in a specific exception dictionary. Unfortunately, it is clear that there is no way to build a database of all proper names in the world.

3.3.3 Prosody

Finding correct intonation, stress, and duration from written text is probably the most challenging problem for years to come. These features together are called prosodic or supra segmental features and may be considered as the melody, rhythm, and emphasis of the speech at the perceptual level. The intonation means how the pitch pattern or fundamental frequency changes during speech. The prosody of continuous speech depends on many separate aspects, such as the meaning of the sentence and the speaker characteristics and emotions. The prosodic dependencies are shown in Figure 3.4. Unfortunately, written text usually contains very little information of these features and some of them change dynamically during speech. However, with some specific control characters this information may be given to a speech synthesizer. Timing at sentence level or grouping of words into phrases correctly is difficult because prosodic phrasing is not always marked in text by punctuation, and phrasal accentuation is almost never marked. If there is no breath pauses in speech or if they are in wrong places, the speech may sound very unnatural or even the meaning of the sentence may be misunderstood.

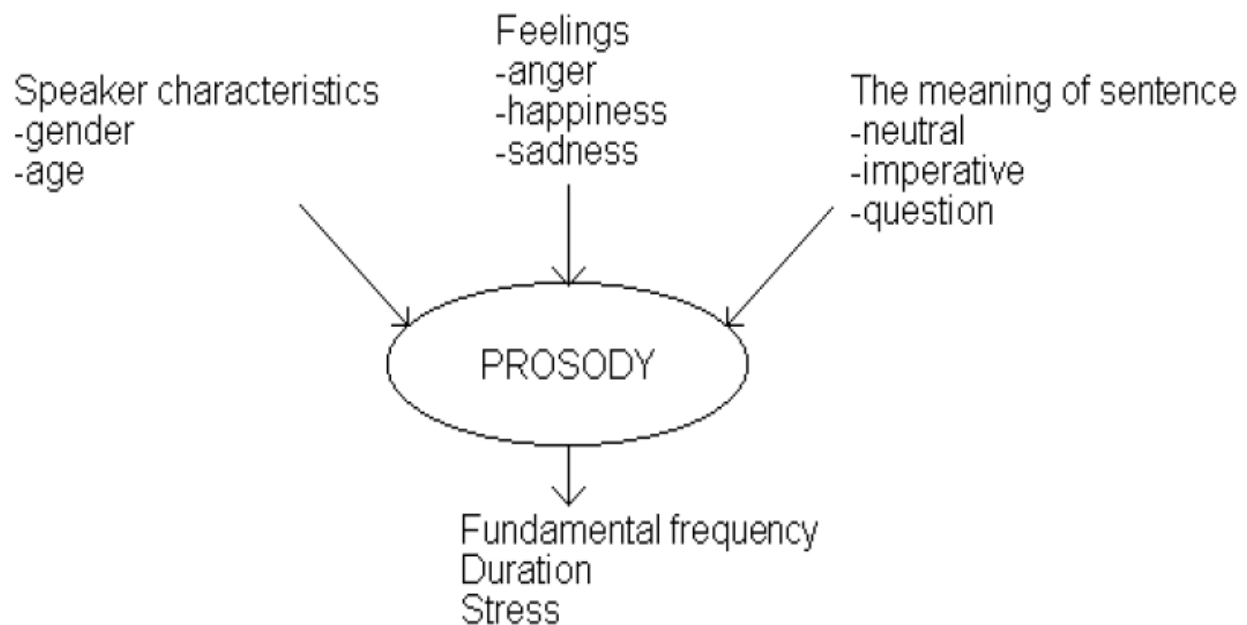


Figure 3.4: Prosodic Dependencies(Speech Processing)

3.4 Speech Synthesized

Synthesized speech can also be manipulated afterwards with normal speech processing algorithms. For example, adding some echo may produce more pleasant speech. However, this

approach may easily increase the computational load of the system. Some experiments to show the use of a combination of the basic synthesis methods have been made, because different methods show different success in generating individual phonemes. Time domain synthesis can produce high-quality and natural sounding speech segments, but in some segment combinations the synthesized speech is discontinuous at the segment boundaries and if a wide-range variation of fundamental frequency is required, overall complexity will increase. On the other hand, formant synthesis yields more homogeneous speech allowing a good control of fundamental frequency, but the voice-timbre sounds more synthetic. This approach leads to the hybrid system which combines the time- and frequency-domain methods. The basic idea of a hybrid system is shown in Figure 3.5.

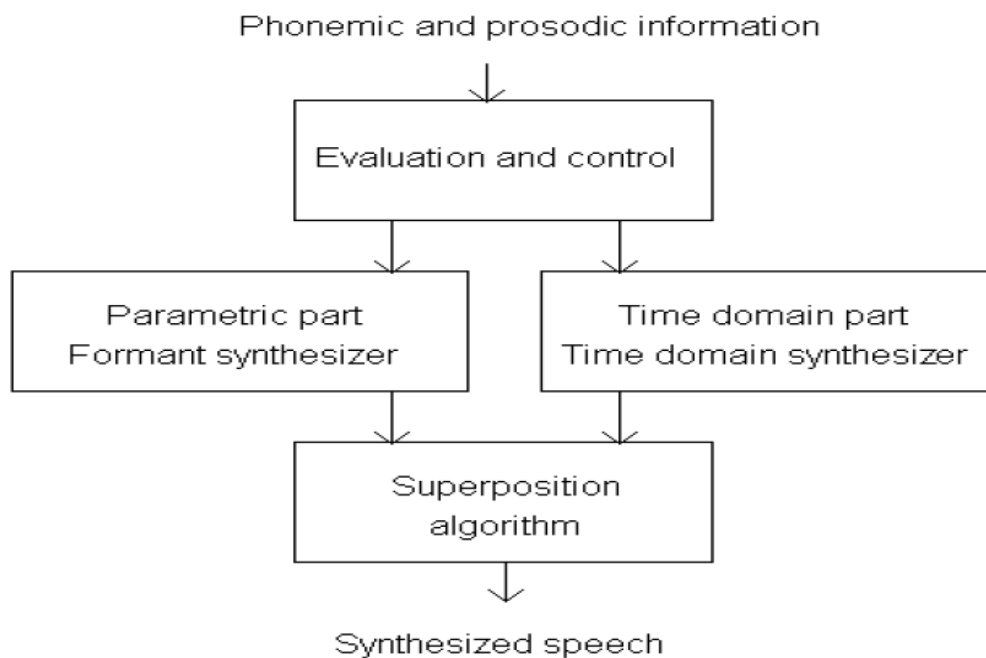


Figure 3.5: Basic idea of the hybrid synthesis system

There are six major classes of spectral analysis algorithms used in speech recognition systems today. The procedures for generating these analyses are summarized in Figure 3.6. Filter bank methods (implemented in analog circuits) were historically the first methods introduced. Linear prediction methods were introduced in the 1970's, and were the dominant technique through the early 1980's. Currently, both Fourier transform and linear prediction techniques

enjoy widespread use in various speech processing applications. In this section, we will discuss each of these techniques, beginning with the digital filter bank.

1) Digital Filter Bank: The digital filter bank is one of the most fundamental concepts in speech processing. A filter bank can be regarded as a crude model of the initial stages of transduction in the human auditory system. There are two main motivations for the filter bank representation. First, the position of maximum displacement along the basilar membrane for stimuli such as pure tones is proportional to the logarithm of the frequency of the tone. This hypothesis is part of a theory of hearing called the “place theory”.

Second, experiments in human perception have shown that frequencies of a complex sound within a certain bandwidth of some nominal frequency cannot be individually identified. When one of the components of this sound falls outside this bandwidth, it can be individually distinguished. We refer to this bandwidth as the critical bandwidth. A critical bandwidth is nominally 10% to 20% of the center frequency of the sound

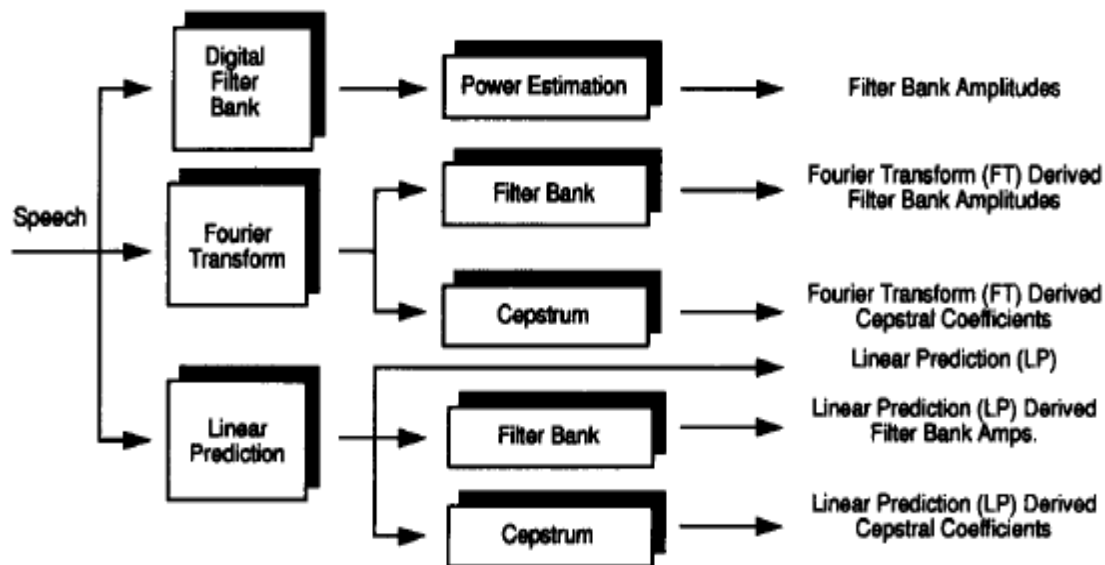


Figure 3.6: The six major spectral analysis algorithms are shown

Cepstral parameters derived from either a Fourier transform or linear prediction model are by far the most popular of these approaches. The Fourier transform methods have traditionally been considered robust in severely noisy environments, and are popular for their similarity to the initial stages of the human auditory system.

We can define a mapping of acoustic frequency f to a "perceptual" frequency scale as follows.

$$\text{Bark} = 13 \tan\left(\frac{0.76 f}{1000}\right) + 3.5 \tan\left(\frac{f^2}{(7600)^2}\right) \quad (3)$$

The units of this perceptual frequency scale are referred as critical band rate, or Bark. The Bark scale is shown in Figure 3.7(a).

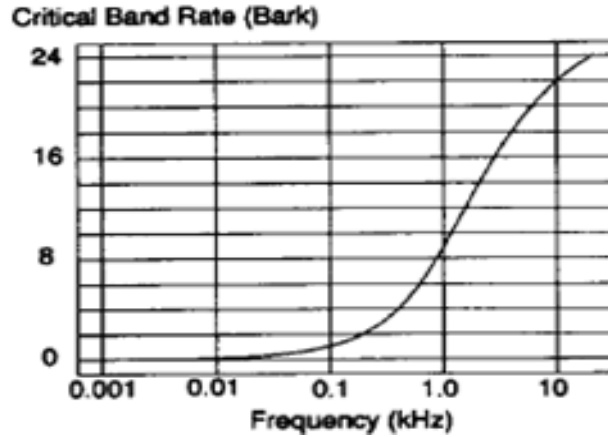


Figure 3.7(a): Bark scale

A more popular approximation to this type of mapping in speech recognition is known as the mel scale.

$$\text{Mel Frequency} = 2595 \log_{10}(1 + f/700) \quad (4)$$

The mel scale attempts to map the perceived frequency of a tone, or pitch, onto a linear scale. This scale is displayed in Figure 3.7(b). It is often approximated as a linear scale from 0 to 1000 Hz, and then a logarithmic scale beyond 1000 Hz.

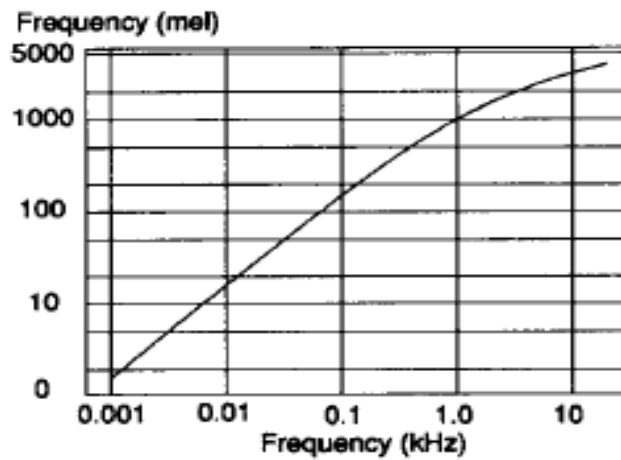


Figure 3.7(b): Mel scale

An expression for critical bandwidth is:

$$BW_{\text{critical}} = 25 + 75[1 + 1.4(f/1000)^2]^{0.69} \quad (5)$$

This transformation can be used to compute bandwidths on a perceptual scale for filters at a given frequency on Bark or mel scales. The critical bandwidth function is also displayed in Figure 3.7(c).

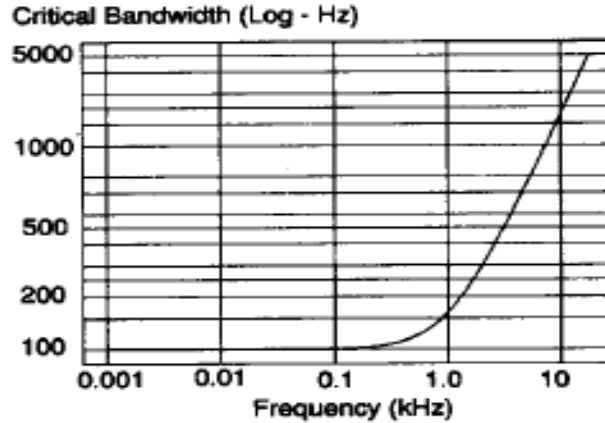


Figure 3.7(c): Critical Band scale

Both the Bark scale and the mel scale can be regarded as a transformation of the frequency scale into a perceptually meaningful scale that is linear. The combination of these two theories gave rise to an analysis technique known as the critical band filter bank.

A critical band filter bank is simply a bank of linear phase FIR band pass filters that are arranged linearly along the Bark (or mel) scale. The bandwidths are chosen to be equal to a critical bandwidth for the corresponding center frequency.

In many speech processing applications, the first filter is omitted, because its range is beyond the capabilities of the AP converter. Often the sampled data collected in the corresponding frequency range are extremely noisy. Telephone grade speech is often processed using a filter bank consisting of 16 bands.

Another equally important filter bank in the speech recognition literature is a filter bank based on the mel scale. In this design, ten filters are assigned linearly from 100 to 1000 Hz. Above 1000 Hz, five filters are assigned for each doubling of the frequency scale (octave). These filters are spaced logarithmically (equally spaced on a log scale). The bandwidths are assigned

such that the 3-dB point is halfway between the current bin and the previous/next bin. The shaded entries in the table are shown only for comparison purposes. Normally, the twenty five samples from the filter bank are used.

Each filter in the digital filter bank is usually implemented as a linear phase filter so that the group delay for all filters is equal to zero, and the output signals from the filters will be synchronized in time. The filter equations for a linear phase filter implementation can be summarized as follows

$$S_i(n) = \sum_{j=-(N_{fbi}-1)/2}^{(N_{fbi}-1)/2} \alpha_{fbi}(j)\delta(n+j) \quad (6)$$

Where $\alpha_{fbi}(j)$ denotes the j th coefficient for the i th critical band filter. The filter order is normally odd for a linear phase filter.

Processing of speech through two filters in this type of filter bank is demonstrated. A speech signal is shown in chapter 4 along with the outputs from two band pass filters, one centered at 250 Hz, and one centered at 2500 Hz. Note that the power of the outputs varies depending on the type of sound spoken. This is the basic merit of the filter bank: certain filter outputs can be correlated with certain classes of speech sounds. The filter outputs are normally processed using any of the power estimation methods previously discussed. Recursive-in-time computations are particularly convenient for post processing in these applications.

The output of this analysis is a vector of power values (or power/frequency pairs) for each frame of data. These are usually combined with other parameters, such as total power, to form the signal measurement vector. The filter bank attempts to decompose the signal into a discrete set of spectral samples that contain information similar to what is presented to higher levels of processing in the auditory system. Because the analysis is largely based entirely on linear processing, the technique is generally robust to ambient noise. We conclude this section with one historical note. Long before computer hardware was capable of performing complex mathematical operations in real time, analog filter banks similar to the digital filter bank previously discussed were used in speech recognition. The filter banks were often built from discrete components and needed to be carefully tuned by adjusting resistors and capacitors. At that time, researchers dreamed of the days when speech recognition system parameters could be

adjusted from software. The analog filter bank is one of the oldest approaches used in speech recognition. Ironically, the analog filter technique has generated some of the lowest cost implementations of speech recognizers to date.

2) The Fourier Transform Filter Bank: We have previously discussed the advantages in using non uniformly spaced frequency samples. One of the easiest and most efficient ways to compute a non uniformly spaced filter bank model of the signal is to simply perform a Fourier transform on the signal, and sample the transform output at the desired frequencies. The Discrete Fourier Transform (DFT) of a signal is defined as

$$S(f) = \sum_{n=0}^{N_s-1} S(n)e^{-j/(2\pi f/f_s)n} \quad (7)$$

Here f denotes the frequency in hertz, $S(f)$ denotes the signal sample frequency and N denotes the window duration in samples.

The filter bank can be implemented on sample the spectrum at the frequencies listed. However, often the spectrum is oversampled at a finer resolution than that described, and each output of the filter bank (a power spectral magnitude) is computed as a weighted sum of its adjacent values

$$S_{avg}(f) = \frac{1}{N_{os}} \sum_{n=0}^{N_{os}} Wfb(n)S(f + \delta f(f, n)) \quad (8)$$

Where N_{os} represents the number of samples used to obtain the averaged value, $Wfb(n)$ represents a weighting function, and $\delta f(f, n)$ represents some function that describes the frequencies in the neighborhood of f to be used in computing the average. Note that the averaging method presented is just one particular method of implementing a spectral smoothing function. Averaging is often performed in the mel scale frequency domain if a DFT is used (since the added computational burden is minimal).[6] Averaging also is usually performed in the log domain (log power values) rather than on spectral amplitudes. The benefit of using averaged values for spectral analysis is demonstrated.

A Fast Fourier Transform (FFT) also can be used as an alternate method of computing the spectrum of the signal. [6] The FFT is a computationally efficient implementation of the DFT under the constraint that the spectrum is to be evaluated at a discrete set of frequencies that are

multiples of f/N . These frequencies are referred to as orthogonal frequencies. The principal advantage of the FFT is that it is very fast: approximately $N \log N$ additions and $N \log N/2$ multiplications are required. (The DFT requires on the order of N^2 operations.) The principal disadvantage is that nonlinear frequency mappings, such as the filter bank.

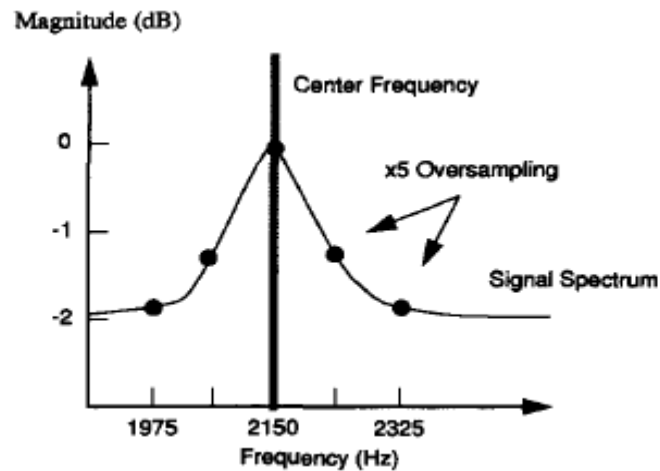


Figure 3.8: Spectrum shown

One additional processing step is often added. Based in part on our sketchy knowledge of human perception, we hypothesize that high-amplitude areas of the spectrum are weighted more heavily in the auditory system than low-amplitude regions. In noisy environments, noise often disproportionately degrades our estimates of the low amplitude areas of the spectrum. Stated another way, we are more confident of the reliability (and repeatability) of our estimates of the high-amplitude areas of the spectrum. For this reason, we often impose a limit on the dynamic range of the spectrum. We refer to this lower limit as the dynamic range threshold. Rather than use noisy estimates of low-amplitude regions of the spectrum, we simply clip, or discard, estimates below a certain threshold from the peak in the spectrum. For Fourier transform-based techniques, this is straightforward to implement as a thresholding function on the spectral magnitude.

3) Linear Prediction Coefficients: We now turn from Fourier transform methods based on linear spectral analysis to a class of parametric modeling techniques that attempt to optimally model the spectrum as an autoregressive process. It is difficult to overstate the impact parametric models have made on speech processing. In last 40 year back, almost every speech processing

system used some sort of algorithm that parametrically fitted the spectrum, whether for recognition, compression, or verification applications. Though parametric models today are less popular in recognition, they are still widely used in compression systems. Parametric models were the impetus for a transition to vastly more powerful statistical modeling techniques in speech recognition. In this section, we will discuss computation of a parametric model based on least mean squared error theory. This technique is known as linear prediction (LP). [10]

The roots of linear prediction, as a least mean squared error algorithm can be traced to many diverse areas: system identification problems in modem control systems, time series analysis for economic applications, maximum entropy techniques, quantum physics, geophysics, adaptive filtering, and spectral estimation in signal processing. Linear prediction theory is well-documented in the literature. Here, we will (very) briefly review the mechanics of computing a linear prediction model, and then discuss the implications in speech recognition. Given a signal $S(n)$ we seek to model the signal as a linear combination of its previous samples. Let us define our signal model as

$$S(n) = - \sum_{i=0}^{Nlp} \alpha_{lp}(i) S(n-i) + e(n) \quad (9)$$

Where Nlp represents the number of coefficients in the model (the order of the predictor), $S(n-i)$ are defined as the linear prediction coefficients (predictor coefficients), and $e(n)$ represents the error in the model (the difference between the predicted value and the actual measured value). [1] One obvious virtue of this model is that, if it is accurate, we should be able to predict future values of the signal based on our current set of measurement. The error term should tell us something about the quality of our model (if the error is small, the model is accurate). It is also possible to show that a linear prediction model effectively models the spectrum of the signal as a smooth spectrum. Equation (9) can be rewritten in 2-transform notation and shown to be a linear filtering operation

$$E(z) = H_{lp} \cdot S(z) \quad (10)$$

Where $E(z)$ and $S(z)$ are the 2-transforms of the error signal and the speech signal, respectively, and

$$H_{lp}(z) = 1 + \sum_{i=0}^{Nlp} \alpha_{lp}(i) z^{-i} \quad (11)$$

Where $\alpha_{lp}(i) = 0$, $H_{lp}(z)$ is defined as the linear prediction inverse filter.

Under the constraint that we would like the mean-squared error to be as small as possible (seeking a solution that gives us the minimum error energy is reasonable), the coefficients of (11) can be obtained from the following matrix equation:

$$\alpha_{lp} = \Phi^{-1} \Phi$$

And

$$\Phi_{n(j,k)} = \frac{1}{N_s} \sum_{m=0}^{N_s-1} S(n+m+j)S(n+m-j) \quad (12)$$

The solution presented in (12) is known as the Covariance Method. α is referred to as a covariance matrix, and $\Phi_{n(j,k)}$ is referred to as the covariance function for $S(n)$. There are three basic ways to compute predictor coefficients: covariance methods based on the covariance matrix (also known as pure least squares methods), autocorrelation methods, and lattice (or harmonic) methods. In speech recognition, the autocorrelation method is almost exclusively used because of its computational efficiency and inherent stability. The autocorrelation method always produces a prediction filter whose zeroes lie inside the unit circle in the z -plane.

The frequency spectrum of a time-domain signal is a representation of that signal in the frequency domain. The frequency spectrum can be generated via a Fourier transform of the signal, and the resulting values are usually presented as amplitude and phase, both plotted versus frequency. Any signal that can be represented as amplitude that varies with time has a corresponding frequency spectrum. This includes familiar concepts such as visible light (color), musical notes, radio/TV channels, and even the regular rotation of the earth. When these physical phenomena are represented in the form of a frequency spectrum, certain physical descriptions of their internal processes become much simpler. Often, the frequency spectrum clearly shows harmonics, visible as distinct spikes or lines, that provide insight into the mechanisms that generate the entire signal. A source of sound can have many different frequencies mixed together. [9] A musical tone's timbre is characterized by its harmonic spectrum. Sound in our environment that we refer to as noise includes many different frequencies. When the frequency spectrum of a sound signal is flat, it is called white noise also known as harmonics.

4.1 Signal Acquisition

- Record the word by recorder which are in audio format
- Convert the audio format to wave format by use the converter say wave file
- Read the wave file in MATLAB
- This is the data size of wave file, used for analysis
- SpectraPlus also used for analysis on wave file

4.2 Spectrum Analysis

Spectrum analysis is the technical process of decomposing a complex signal into simpler parts. As described above, many physical processes are best described as a sum of many individual frequency components. Any process that quantifies the various amounts (e.g. amplitudes, powers, intensities, or phases), versus frequency can be called spectrum analysis. Spectrum analysis can be performed on the entire signal. Alternatively, a signal can be broken into short segments (sometimes called frames), and spectrum analysis may be applied to these individual segments. Periodic functions (such as $\sin(t)$) are particularly well-suited for this sub-

division. General mathematical techniques for analyzing non-periodic functions fall into the category of Fourier analysis. The Fourier transform of a function produces a frequency spectrum which contains all of the information about the original signal, but in a different form. This means that the original function can be completely reconstructed (synthesized) by an inverse Fourier transform. For perfect reconstruction, the spectrum analyzer must preserve both the amplitude and phase of each frequency component. These two pieces of information can be represented as a 2-dimensional vector, as a complex number, or as magnitude (amplitude) and phase in polar coordinates. A common technique in signal processing is to consider the squared amplitude, or power; in this case the resulting plot is referred to as a power spectrum. In practice, nearly all software and electronic devices that generate frequency spectra apply a fast Fourier transform (FFT), which is a specific mathematical approximation to the full integral solution. Formally stated, the FFT is a method for computing the discrete Fourier transform of a sampled signal. Because of reversibility, the Fourier transform is called a representation of the function, in terms of frequency instead of time; thus, it is a frequency domain representation.[10] Linear operations that could be performed in the time domain have counterparts that can often be performed more easily in the frequency domain. Frequency analysis also simplifies the understanding and interpretation of the effects of various time-domain operations, both linear and non-linear. For instance, only non-linear operations can create new frequencies in the frequency spectrum. The Fourier transform of a stochastic (random) waveform (noise) is also random. Some kind of averaging is required in order to create a clear picture of the underlying frequency content (frequency distribution). Typically, the data is divided into time-segments of a chosen duration, and transforms are performed on each one. Then the magnitude or (usually) squared-magnitude components of the transforms are summed into an average transform. This is a very common operation performed on digitally sampled time-domain data, using the discrete Fourier transform. This type of processing is called Welch's method.

Here, we recognition the speech word using spectrum analysis. We record twenty five words which are different from each other. We record one word at five times and then analysis at five times on that word which give us spectrum of signal. By analysis the word spectrum we find various parameters of word which like's Mean, Median, Standard Deviation(STD), Root Mean Square(RMS), Maximum Peak, Minimum Peak, Slope of maximum peak, Width of maximum peak, Signal to Noise Ratio(SNR), Peak Frequency, Peak Amplitude, Total Power, Total

Harmonic Distortion(THD), THD + Noise, Inter Modulation Distortion(IMD) using MATLAB and SpectraPlus. When we find these parameters we get various values in which these parameters are lying. These parameters have some values in which they are bounded. Out of these bound values they are refer to another word parameter show.

To recognition the speech word, we record the word and convert into wave format. And store in MATLAB dictionary. We are work on the speech signal so we use Digital Signal Processing. It is necessary to convert our samples in a series of data that we can interpret. To do this, we record our samples of sound in an archive with “.wav” extension, and we retrieve it using the instruction “wavread” in MATLAB. To represent the signal in the frequency domain we will use the DFT. A way to apply the DFT, is using the FFT. The condition to apply the FFT is that the length of our signal must be a power of two. Next we will get the FFT of our signal. The real and imaginary components of the FFT of the signal are stored in the vector x where x is read the file name. Finally mag represents the magnitude of the FFT (x) or the absolute of FFT which is show in figure 4.1 and spectrum form shown in figure 4.2 (b). The basic algorithm is;

- Read the sound file say (x)
- Implementation of the FFT to represent the signal in the dominion of the frequency say(y).
- Take the absolute of y say (z).
- Plot(z)

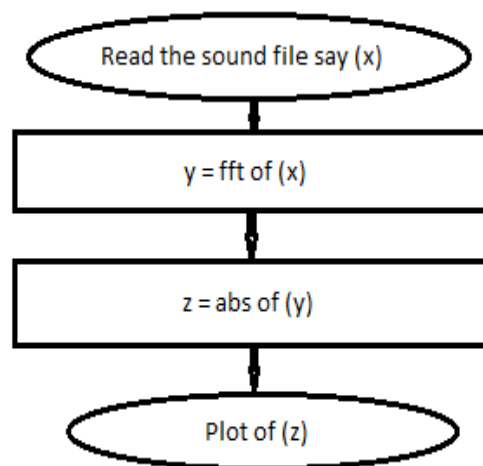


Figure 4.1: Flow-Chat of algorithm

The result of this algorithm with the following parts:

- a) Voice Signal - This is the signal of word. Figure 4.2 (a)
- b) Spectrum - This is the spectrum of signal by using algorithm. Figure 4.2 (b)

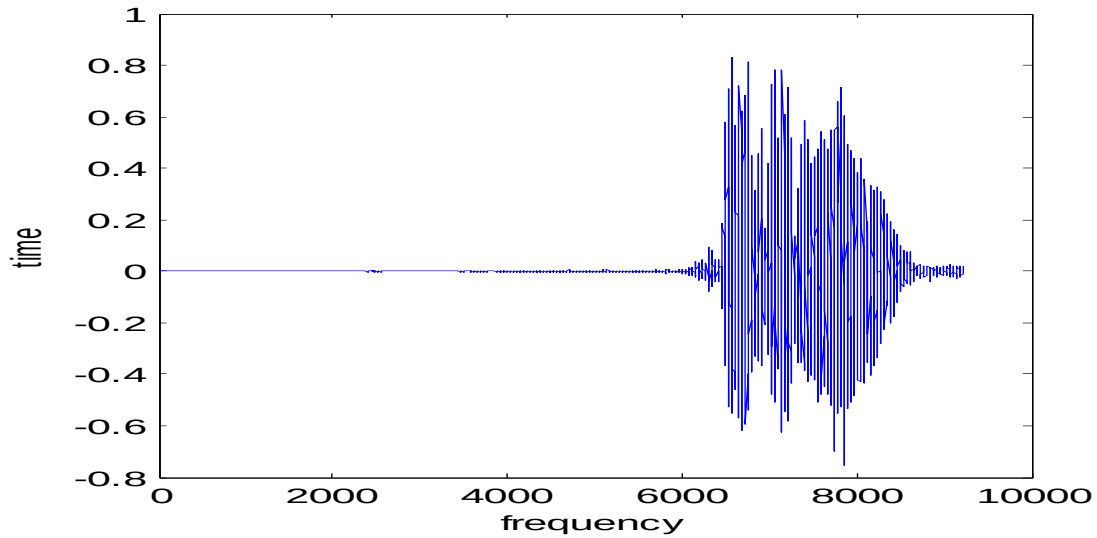


Figure 4.2 (a): Signal of word

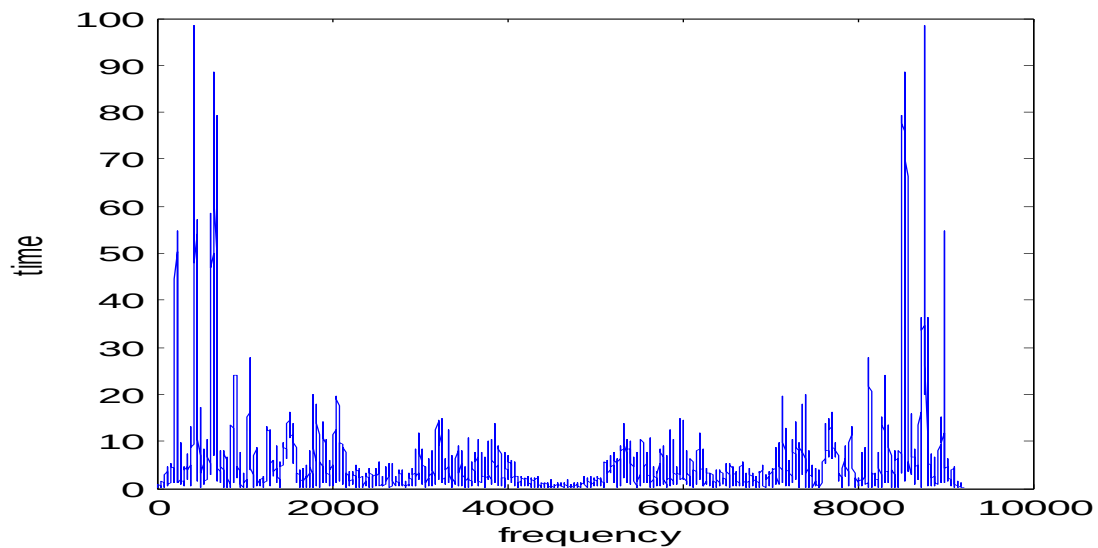


Figure 4.2 (b): Spectrum of signal

Next we will show the algorithm to draw the representation of our file of sound in spectrum form and then analysis on this spectrum to get various parameters. After above Algorithm, plotting the spectrum and we are find parameters by this algorithm.

- Maximum peak of z
- Minimum peak of z
- Find the slop of maximum peak of z by mathematic method
- Find the width of maximum slop peak
- Root Mean Square(RMS) of z
- Mean of z
- Median of z
- Standard Deviation of z

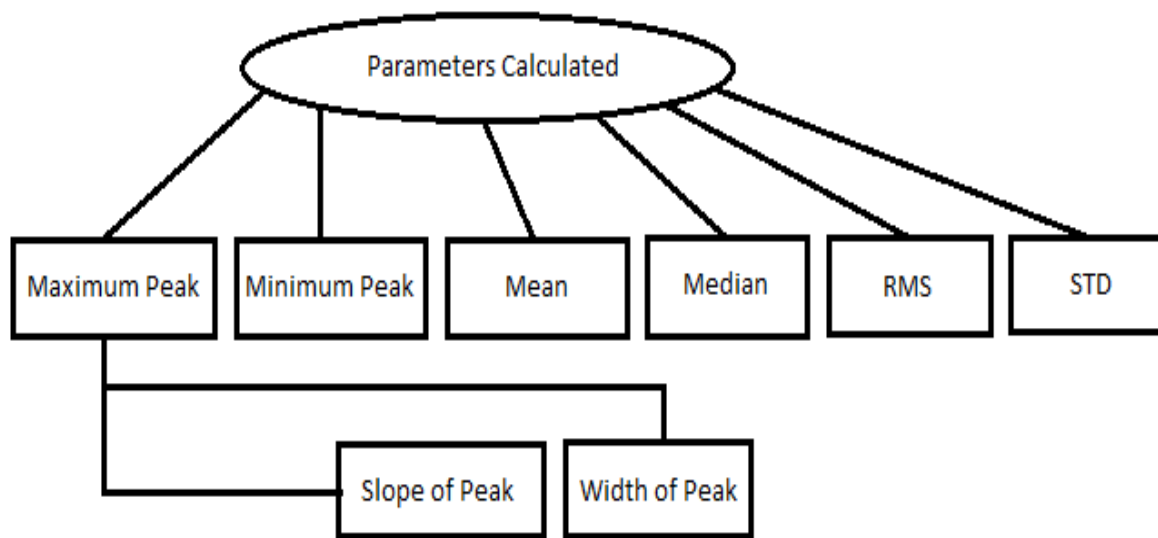


Figure 4.3: Parameters used in algorithm

By analysis the signal spectrum we get some parameters (Mean, Median, Standard Deviation, Root Mean Square (RMS), Maximum Peak, Minimum Peak, Slop of maximum Peak, Width of maximum peak, Signal to Noise Ratio(SNR), Peak Frequency, Peak Amplitude, Total Power, Total Harmonic Distortion(THD), THD + Noise, Inter Modulation Distortion(IMD)). These parameters details as;

4.3 MATLAB Parameters

Mean

The mean may often be confused with the median, mode or range. The mean is the arithmetic average of a set of values, or distribution; however, for skewed distributions, the mean

is not necessarily the same as the middle value (median), or the most likely (mode). For example, mean income is skewed upwards by a small number of people with very large incomes, so that the majority has an income lower than the mean. By contrast, the median income is the level at which half the population is below and half is above. The mode income is the most likely income, and favors the larger number of people with lower incomes. The median or mode are often more intuitive measures of such data. In MATLAB, we get it by this command;

```
>> mean(z);
```

Median

The median is the middle number. The median of a finite list of numbers can be found by arranging all the observations from lowest value to highest value and picking the middle one. If there is an even number of observations, then there is no single middle value; the median is then usually defined to be the mean of the two middle values. In a sample of data, or a finite population, there may be no member of the sample whose value is identical to the median (in the case of an even sample size), and, if there is such a member, there may be more than one so that the median may not uniquely identify a sample member. Nonetheless, the value of the median is uniquely determined with the usual definition. In MATLAB, we get it by this command;

```
>>median(z);
```

Standard Deviation

Standard deviation is a widely used measurement of variability or diversity used in statistics and probability theory. It shows how much variation or "dispersion" there is from the average (mean, or expected value). A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data are spread out over a large range of values. Technically, the standard deviation of a statistical population, data set, or probability distribution is the square root of its variance. It is algebraically simpler though practically less robust than the average absolute deviation. A useful property of standard deviation is that, unlike variance, it is expressed in the same units as the data. In addition to expressing the variability of a population, standard deviation is commonly used to measure confidence in statistical conclusions. In science, researchers commonly report the standard deviation of experimental data, and only effects that fall far outside the range of standard

deviation are considered statistically significant – normal random error or variation in the measurements is in this way distinguished from causal variation. In MATLAB, we get it by this command;

```
>> std(z);
```

Root Mean Square (RMS)

The RMS value of a set of values (or a continuous-time waveform) is the square root of the arithmetic mean (average) of the squares of the original values (or the square of the function that defines the continuous waveform). In mathematics, the root mean square (abbreviated RMS or rms), also known as the quadratic mean, is a statistical measure of the magnitude of a varying quantity. It is especially useful when variants are positive and negative, e.g., sinusoids. RMS is used in various fields, including electrical engineering; one of the more prominent uses of RMS is in the field of signal amplifiers. It can be calculated for a series of discrete values or for a continuously varying function. The name comes from the fact that it is the square root of the mean of the squares of the values. The RMS over all time of a periodic function is equal to the RMS of one period of the function. The RMS value of a continuous function or signal can be approximated by taking the RMS of a series of equally spaced samples. Additionally, the RMS value of various waveforms can also be determined without calculus. In MATLAB, we get it by this command;

```
>> n = length(z);
```

```
>> norm(z);
```

```
>> rms = norm(z)/ sqrt(n);
```

Maximum Peak

Maximum peak is the height of peak in which highest value are lying. In signal spectrum we find the maximum peak which tells us the one important parameter of that signal. In figure 4.4 shows the maximum peak. In MATLAB, we get it by this command;

```
>>max(z);
```

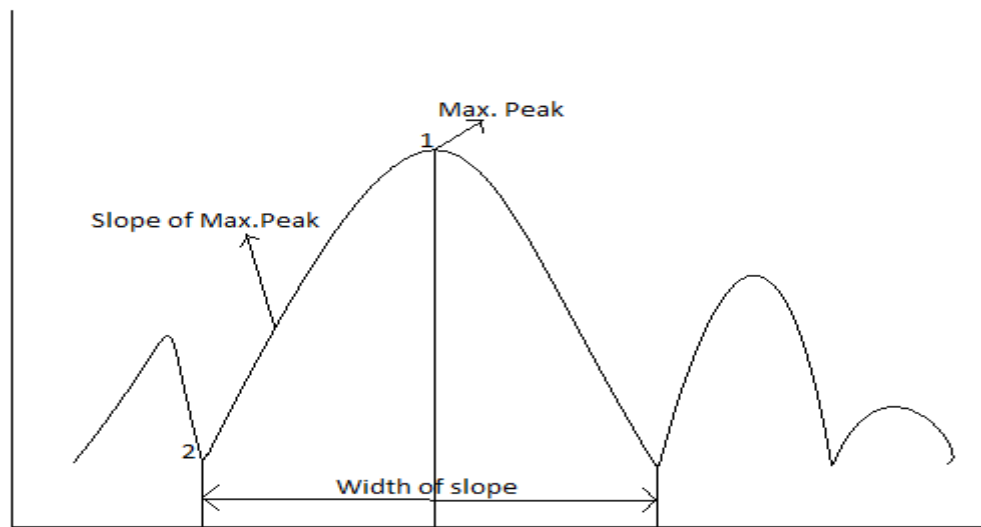


Figure 4.4: Parameters shown

Minimum Peak

Minimum peak is the lower height of peak in which lowest value is lying. In signal spectrum we find the minimum peak which tells us the one important parameter of that signal. In figure 4.4 shows the minimum peak. In MATLAB, we get it by this command;

```
>>min(z);
```

Width of Peak

This is measured by drawing tangents to the front and back slopes of the peak. The baseline width is the distance between the intersections of these tangents with the baseline. The width is the distance from the front slope of the peak to the back slope of the peak measured at 50% of the maximum peak height. In figure4.4 show the width of peak.

$$\text{Width of peak} = (x_2 - x_1) \quad (1)$$

Slope of Peak

In mathematics the slope or gradient of a line describes its steepness, incline, or grade. A higher slope value indicates a steeper incline. The slope is (in the simplest of terms) the measurement of a line, and is defined as the ratio of the "rise" divided by the "run" between two points on a line, or in other words, the ratio of the altitude change to the horizontal distance

between any two points on the line. Given two points (x_1, y_1) and (x_2, y_2) on a line, the slope m of the line is

$$m = \frac{(y_2 - y_1)}{(x_2 - x_1)} \quad (2)$$

4.4 SpectraPlus Parameters

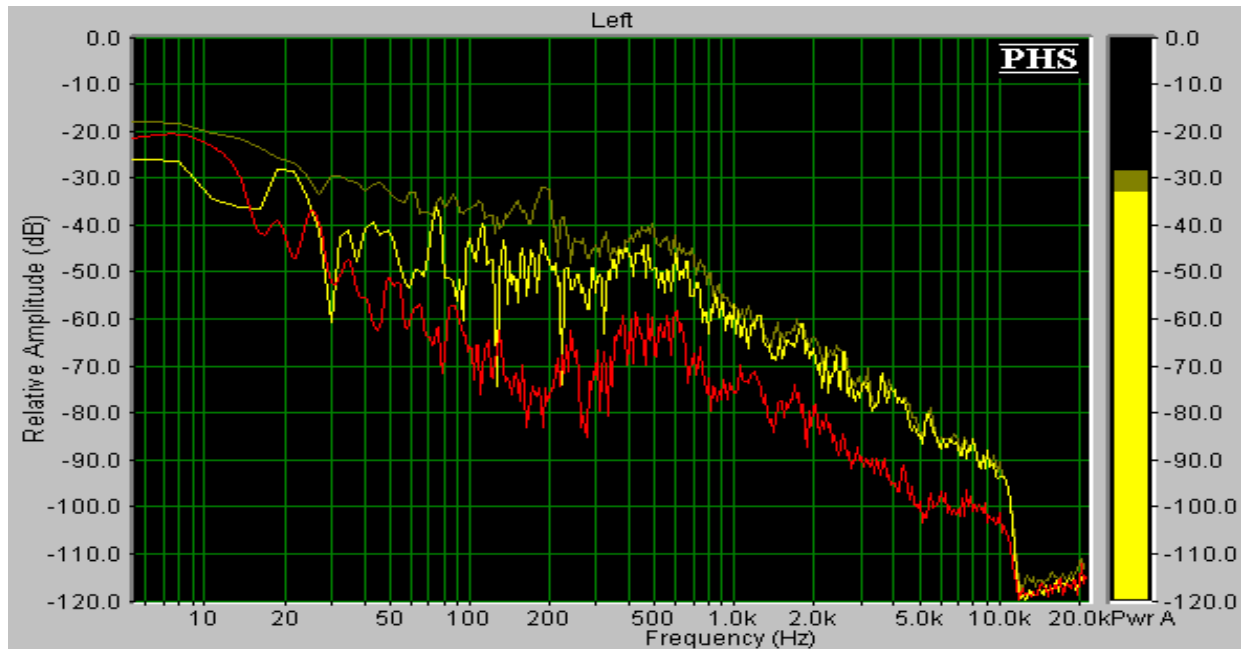


Figure 4.5: Spectrum

Peak Frequency

The peak frequency utility will display the frequency of the strongest spectral component in the entire span (not just what is displayed). This utility window can be resized as required by the user and its contents will update anytime a new spectrum is computed. If the analyzer is configured to display the spectrum of both the left and right channels, this utility window will also show the value for both channels. The left channel value will appear above the right channel value.

Peak Amplitude

The peak amplitude utility will display the amplitude of the strongest spectral component in the entire span (not just what is displayed). This utility window can be resized as required by the user and its contents will update anytime a new spectrum is computed. If the analyzer is

configured to display the spectrum of both the left and right channels, this utility window will also show the value for both channels. The left channel value will appear above the right channel value.

Total Power

The total power utility will display the total RMS power level for the entire spectrum and is equal to the total power bar on the right side of the spectrum view. This is equivalent to the readings on a Sound Level Meter with calibrated to units of SPL. This utility window can be resized as required by the user and its contents will update anytime a new spectrum is computed. If the analyzer is configured to display the spectrum of both the left and right channels, this utility window will also show the value for both channels. The left channel value will appear above the right channel value.

Total Harmonic Distortion

Real world audio devices introduce unwanted side effects called distortion in addition to the desired response. Total Harmonic Distortion (THD) and (THD+Noise) are two widely used measurements of this quantity. Both measurements are expressed as a percentage. The lower the THD value, the "cleaner" the response. Total Harmonic Distortion (THD) is the ratio of the harmonic power to the fundamental power. It is computed by searching the entire spectrum to find the peak frequency (fundamental) and then calculating the total power in the harmonic frequencies. The THD level is then computed as the ratio of the total harmonic power to the fundamental power. Residual noise is not included in this calculation.

Total Harmonic Distortion + Noise (THD+N)

THD + Noise are the ratio of the harmonic power plus noise to the fundamental power. It is computed by searching the entire spectrum to find the peak frequency (fundamental) and then calculating the total power in the remaining spectrum (harmonics plus noise). The THD level is then computed as the ratio of the total harmonic+noise power to the fundamental power. THD+N values will almost always be greater than the THD values for the same signal.

A typical THD measurement will use a 1k Hz reference test tone. This reference signal is passed through the audio device under test and its spectrum computed. It is often useful to employ long term averaging to cancel the random noise components and enhance the

measurement accuracy. It should be noted that the purity of both the sound card and the reference signal must exceed the device being measured if meaningful results are to be expected. The purity of the signal generator is directly dependent on the THD of the sound card used.

This utility window can be resized as required by the user and its contents will update anytime a new spectrum is computed. If the analyzer is configured to display the spectrum of both the left and right channels, this utility window will also show the value for both channels. The left channel value will appear above the right channel value.

Inter Modulation Distortion

Inter Modulation Distortion (IMD) is a measure of the distortion caused by the interaction (mixing) of two tones. When multiple signals are injected into a device, undesired modulation or mixing of these two signals can occur. The IMD level is calculated by first computing the frequencies and amplitudes of the two strongest tones in the spectrum. The total power in each of the inter modulation product frequencies is then computed. IMD is the ratio of the inter modulation power to the rms sum of the tone power.

A typical IMD measurement will use tones at 250 Hz and 8020 Hz with the higher frequency tone reduced by 12dB (4:1 ratio) below the low frequency tone. It is often useful to employ long term averaging to cancel the random noise components and enhance the measurement accuracy. This utility window can be resized as required by the user and its contents will update anytime a new spectrum is computed. If the analyzer is configured to display the spectrum of both the left and right channels, this utility window will also show the value for both channels. The left channel value will appear above the right channel value.

Signal to Noise Ratio

The Signal to Noise Ratio (SNR) is the ratio of the signal peak power level to the total noise level and is expressed in decibels (dB). The SNR is computed by searching the entire spectrum to find the peak frequency and then calculating the total noise power in the remaining spectrum. The SNR is then computed as the ratio of the noise power to the peak power and expressed in decibels. This utility window can be resized as required by the user and its contents will update anytime a new spectrum is computed. If the analyzer is configured to display the

spectrum of both the left and right channels, this utility window will also show the value for both channels. The left channel value will appear above the right channel value.

4.5 LPC Parameters

Speech signals are recorded using an audio wave recorder in the normal room environment. The recorded speech signal is passed through the speech starting to detect the presence of the speech signal and remove the silence and pauses portions of the signals. The resulting signal is then filtered for the removal of unwanted background noise from the speech signals. The filtered signal is then windowed ensuring half frame overlap. After windowing, the speech signal is then subjected to calculate the LPC coefficient.[14] The feature extractor uses a standard LPC Cepstrum coder, which converts the incoming speech signal into LPC Cepstrum coefficients. In MATLAB, by LPC Spectrum programming we get another spectrum which are made from original signal spectrum. In figure 4.6 Red line show the LPC coefficients spectrum and Blue line show the original signal of word. By analysis this Red line spectrum we find another various parameters which are totally different from above parameters. The LPC spectrum algorithm as shown;

- Take the framesize of sample
- Take length of fft sample
- Create a system object to read from a specified multimedia file say (hfileIn)
- Take sampling frequency which is multiple of audio sample rate and multimedia file(hfileIn)
- Create an FIR digital filter system object for pre-emphasis say (hpreemphasis)
- Create a buffer system object say (hbuf)
- Create a window system object say (hwindow)
- Create an auto correlator system object say (hacf)
- Create an levinson system object which give output both polynomial and reflection coefficients say (hlevinson)
- Create an FIR digital filter system object used for analysis say (hanalysis)
- Create two all pole digital filter system object used for synthesis say (hsynthesis)
- Create an de-emphasis system object say (hdeemphasis)
- Create a audio player system object to play the resulting audio say (haudioout)

- Plot the setup
- Make the loop for done file(hfileIn)
- Read the audio input say (sig)
- Now analysis the audio input by various steps which we creates them above
- Analysis Pre-emphasis say (sigpreem)
- Analysis Buffer and Window say (sigwin)
- Analysis Autocorrelation say (sigacf)
- Analysis Levinson-Durbin say (sigA, sigK)
- Analysis the filter say (siglpc)
- Now synthesis the steps
- Synthesis the filter say (sigsyn)
- De-emphasis say (sigout)
- Now play audio output implement with Step
- Plot the lpc data say (s)
- End the loop

By using this algorithm we get the spectrum. In this figure 4.6 Blue line is original signal and Red line show the LPC coefficient of the original signal of the word.

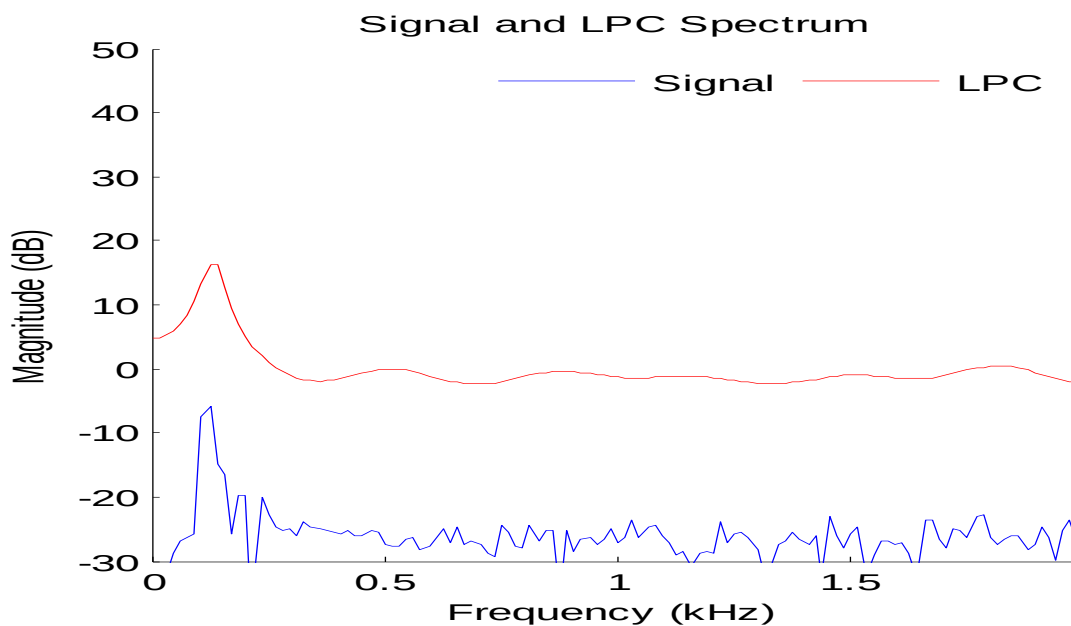


Figure 4.6: LPC spectrum (Red line) and Original signal spectrum (Blue line)

We get the LPC spectrum. To analysis this LPC spectrum, we use workspace folder in MATLAB. In workspace folder we get this LPC spectrum values which are only in row matrices. These row values are given this LPC spectrum, but we get time-frequency domain spectrum shown figure 4.7. Then we collect all these row values and by plotting graphs we get another time-frequency domain spectrum. Now using algorithm which analysis (Statistical Methods) this Time and Frequency domain spectrum and get another various parameters which are Maximum Peak, Minimum Peak, Standard Deviation(STD), Root Mean Square(RMS), Slope of Peak, Width of Peak, Maximum Magnitude at Frequency, Attenuation.

- Maximum peak of LPC spectrum
- Minimum peak of LPC spectrum
- Find the slop of maximum peak of LPC spectrum by mathematic method
- Find the width of slop peak
- Root Mean Square(RMS) of LPC spectrum
- Standard Deviation of LPC spectrum
- Peak Frequency
- Maximum amplitude at peak of frequency
- Attenuation

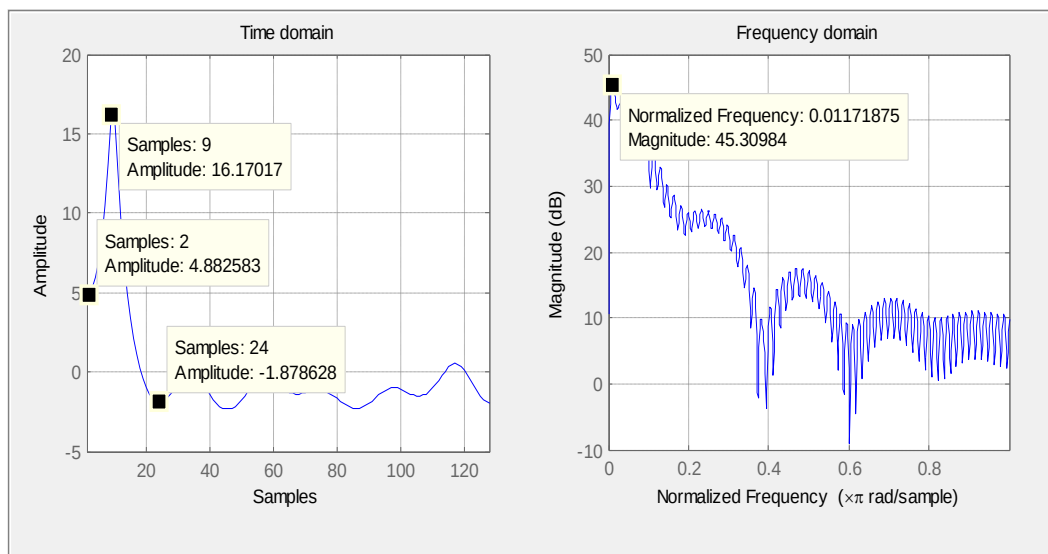


Figure 4.7: Time–Frequency domain spectrum

5.1 Results

To recognition the speech word we record twenty five words in English. These words are Army, Police, Boom, Bomb, Bus, But, Put, Fill, Kill, Singh, King, Kite, Night, Punjab, India, Pakistan, Minister, Murder, England, Canada, America, Australia, Japan, China and Rdx. By using Matlab and SpectraPlus we analysis their signal spectrum and find various parameters as shown in tables. We record the five samples of single word and analysis their five samples spectrum. By analysis the spectrum of every word we find various parameters. At analysis, we get different values of same parameters in single word. Every parameter has their own values in which that are lying. By get these parameters values we shown that each parameter has their range of values. Out of range values, parameters are change and they are lying to another word parameters. Now in this chapter first we are show spectrum of the words and then analysis on them. 2nd we show parameters results of different word in five times records and 3rd from these results we plot some graphs which show their range of values. And at last some discussion on their results and graphs.

In tables parameters are Mean, Median, Standard Deviation(STD), Root Mean Square(RMS), Maximum Peak, Minimum Peak, Slope of maximum peak, Width of maximum peak, Signal to Noise Ratio(SNR), Peak Frequency, Peak Amplitude, Total Power, Total Harmonic Distortion(THD), THD + Noise, Inter Modulation Distortion(IMD), Maximum Magnitude at Frequency, Attenuation.

S1, S2, S3, S4, S5 are the samples of each word.

Avg ~ Average value

Stdev ~ Standard deviation value

In graph (a), we show clear only some parameters, but in graph (b) which show all those parameters which are not clearly show in graph (a). These graph (a) and (b) are show parameters range values from spectrum analysis. And graph (c) is also show the parameters range values from LPC spectrum analysis.

1 ARMY

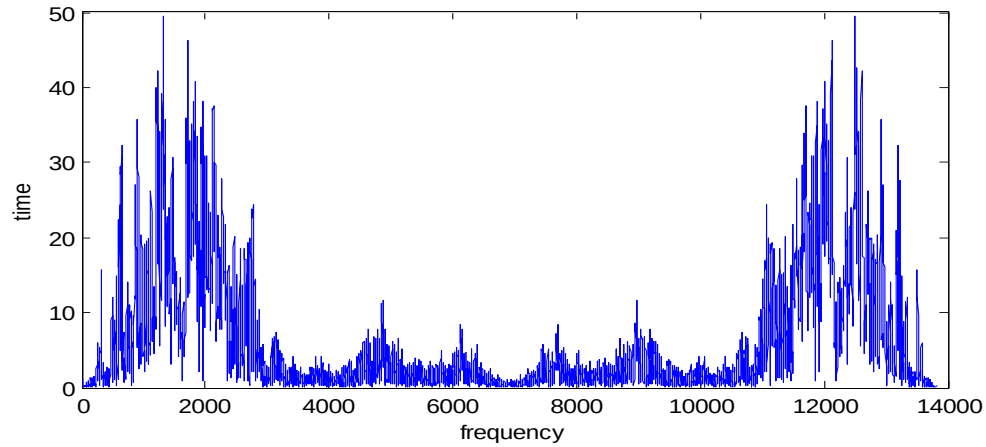


Figure 5.1: Spectrum of 'ARMY' word

By Spectrum Analysis

Table 5.1(a): Parameters of the word 'ARMY'

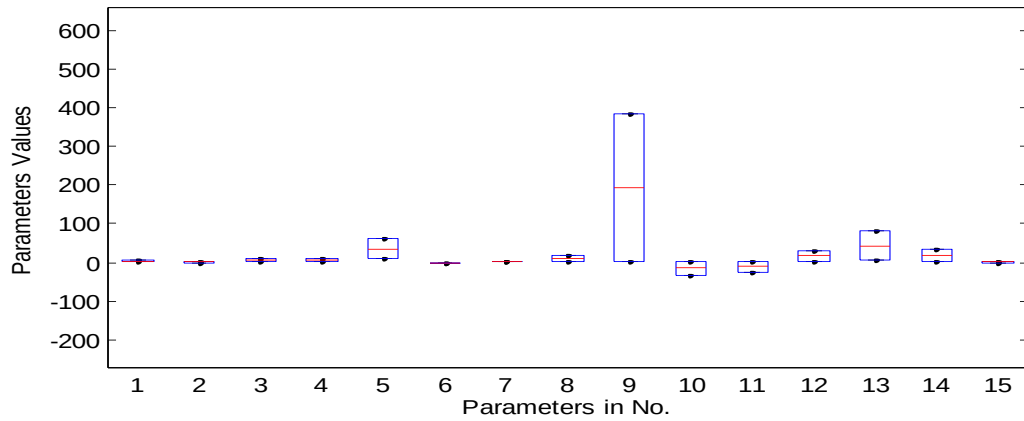
	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	6.3565	6.1544	6.1074	5.2993	6.2731	6.04	0.42
MEDIAN	2.4722	2.8261	2.6161	2.4937	2.3947	2.56	0.17
STD	8.2535	8.1002	9.571	8.8191	10.284	9.01	0.92
RMS	10.4174	10.1727	11.3533	10.2481	12.384	10.92	0.95
MAX	49.399	52.8677	71.5669	66.2983	65.2968	61.09	9.47
MIN	0.0035	0.0116	0.0043	0.004	0.0039	0.01	0.00
SLOP	2.84	4.21	2.46	2.97	3.29	3.15	0.66
WIDTH	16	19	15	17	17	16.80	1.48
P.FREQ(Hz)	380.86	388.67	384.29	385.19	379.92	383.79	3.52
P.AMP(dB)	-29.05	-35.86	-33.84	-30.69	-34.71	-32.83	2.85
T.POW(dB)	-21.19	-26.04	-23.89	-24.71	-22.68	-23.70	1.86
THD(%)	29.38	33.29	30.75	31.74	32.69	31.57	1.56
THD+N (%)	69.88	90.98	78.39	84.19	79.37	80.56	7.78
IMD(%)	30.94	34.42	33.5	29.39	32.98	32.25	2.04
SNR(dB)	3.11	2.95	3.09	3.26	2.89	3.06	0.15

In this table parameters are Mean, Median, Standard Deviation(STD), Root Mean Square(RMS), Maximum Peak, Minimum Peak, Slope of maximum peak, Width of maximum peak, Peak Frequency, Peak Amplitude, Total Power, Total Harmonic Distortion(THD), THD + Noise, Inter

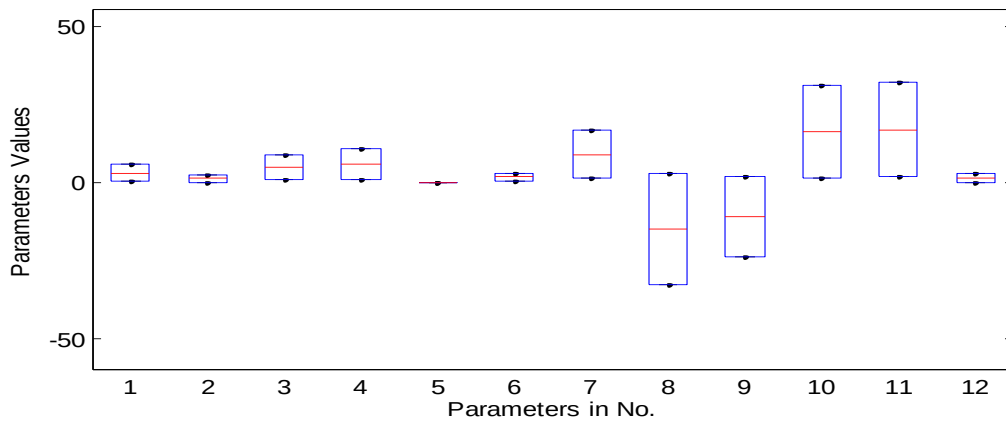
Modulation Distortion(IMD), Signal to Noise Ratio(SNR) and S1, S2, S3, S4, S5 are the samples of the army word.

Avg ~ Average value

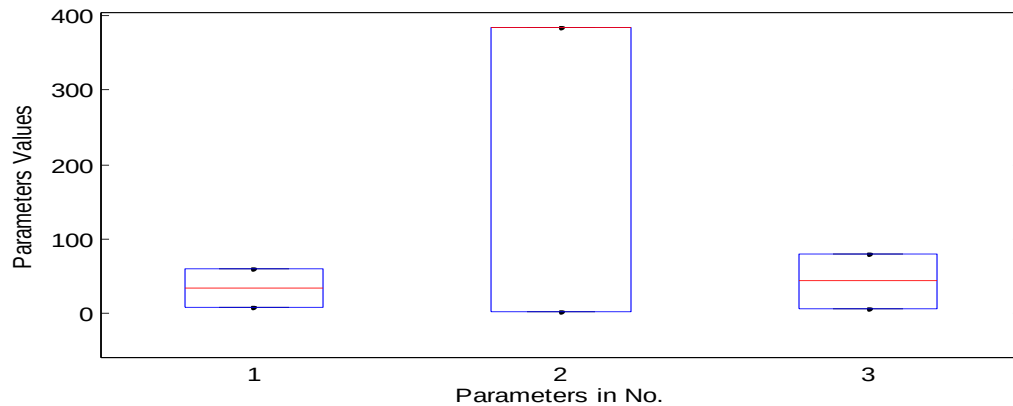
Stdev ~ Standard deviation value



Graph 5.1(a): Parameters show their range of values from table 5.1(a)



Graph 5.1(b): Insignificant parameters from table 5.1(a)



Graph 5.1(c): Significant Parameters from table 5.1(a)

By LPC Spectrum Analysis

Table 5.1(b): Parameters of the word ‘ARMY’

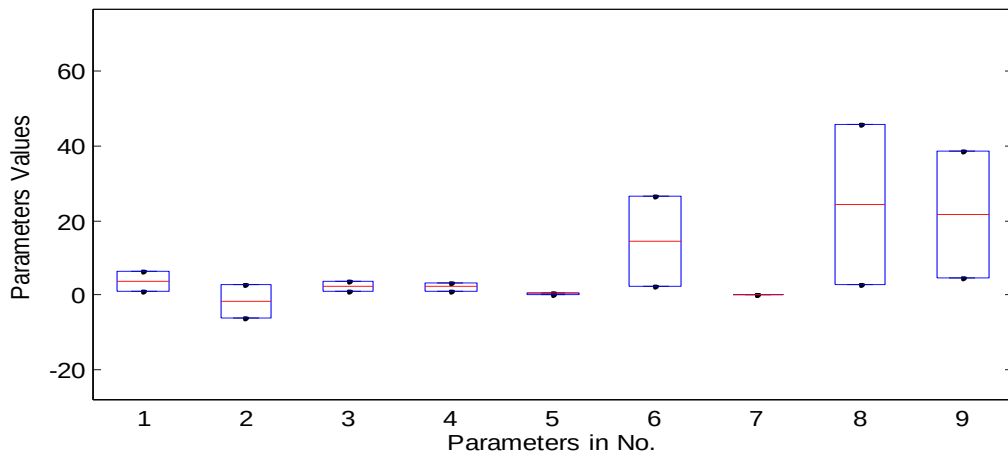
	S1	S2	S3	S4	S5	Avg	Stdev
MAX	4.8034	7.7358	5.8134	6.2934	6.7931	6.29	1.09
MIN	-11.276	-5.8097	-4.0146	-4.834	-5.414	-6.27	2.88
STD	4.5898	3.0454	2.2528	3.7201	3.892	3.50	0.89
RMS	4.5721	3.0336	2.244	3.0594	4.0021	3.38	0.91
SLOP	0.47	0.71	0.61	0.6	0.58	0.59	0.09
WIDTH	26	23	29	27	28	26.60	2.30
FRQ	0.0097	0.031	0.0175	0.0119	0.0127	0.02	0.01
MAXMAG	49.75	44.72	41.88	46.67	45.28	45.66	2.88
A.TION(dB)	33	36.4	44.3	40.8	37.6	38.42	4.31

In this table parameters are Maximum Peak, Minimum Peak, Standard Deviation(STD), Root Mean Square(RMS), Slope of maximum peak, Width of maximum peak, Maximum Magnitude, at Frequency, Attenuation.

S1, S2, S3, S4, S5 are the samples of the army word.

Avg ~ Average value

Stdev ~ Standard deviation value



Graph 5.1(d): Significant Parameters from table 5.1(b)

In the graph 5.1(a) all parameters are shown which not clarify the values so we plot another graphs which are show all values likes in graphs (b) and (c). In the graph 5.1(d) show LPC parameter values of the word ARMY. These graphs show the range of parameter values. Parameters are lying in bounded values which likes shown in these graphs. Similarly next we

plot only those graphs which show all parameter values means significant and insignificant parameters.

2 POLICE

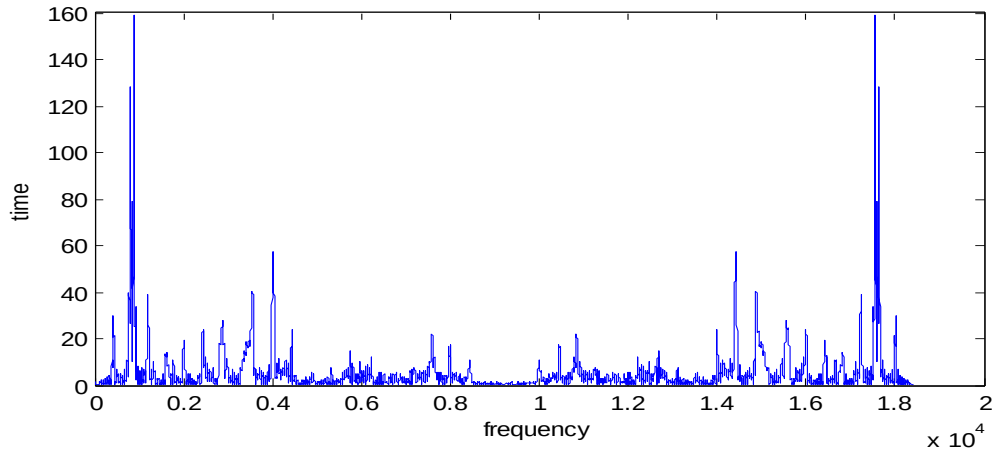
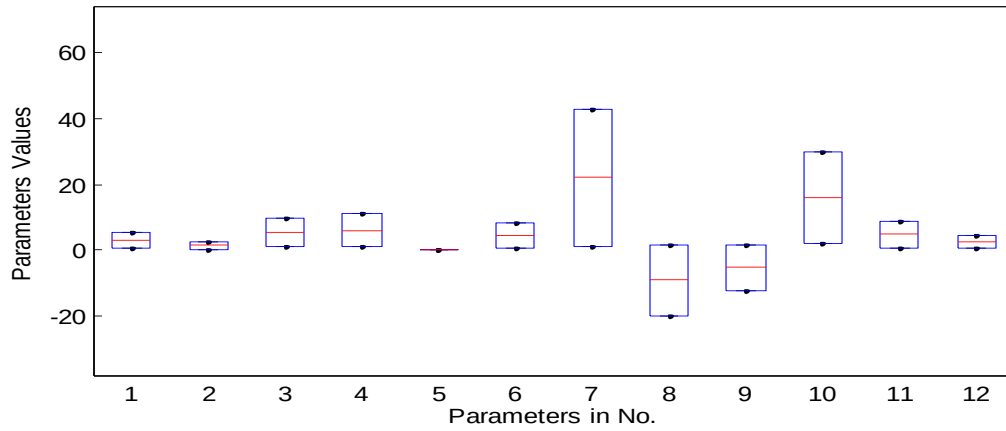


Figure 5.2: Spectrum of 'POLICE' word

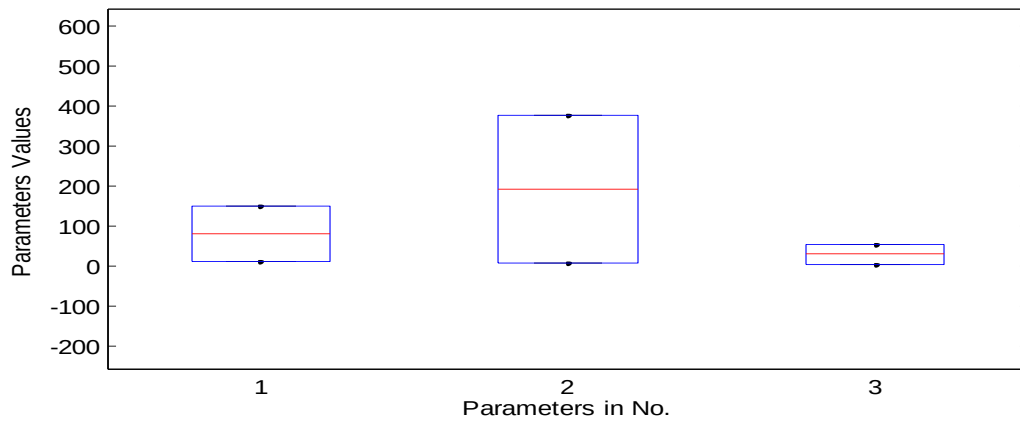
By Spectrum Analysis

Table 5.2(a): Parameters of word 'POLICE'

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	5.4265	5.8427	4.8558	4.479	5.2836	5.18	0.53
MEDIAN	2.9839	2.9468	2.639	2.3981	2.649	2.72	0.24
STD	9.8835	11.178	8.7827	8.3843	9.2783	9.50	1.09
RMS	11.2749	12.6132	10.0355	11.1264	10.1028	11.03	1.05
MAX	159.976	159.12	132.281	148.352	150.283	150.00	11.17
MIN	0.0158	0.0074	0.0068	0.0129	0.0149	0.012	0.004
SLOP	9	7.23	7.67	8.36	8.36	8.12	0.69
WIDTH	44	44	43	42	42	43.00	1.00
P.FREQ(Hz)	359.38	382.81	380.86	379.36	373.87	375.26	9.48
P.AMP(dB)	-21.29	-17.24	-20.4	-19.37	-20.81	-19.82	1.61
T.POW(dB)	-14.52	-9.85	-12.94	-11.43	-12.48	-12.24	1.74
THD(%)	31.73	32.45	27.51	29.64	28.47	29.96	2.10
THD+N (%)	57.97	58.36	50.01	56.38	57.03	55.95	3.41
IMD(%)	9.067	9.56	8.68	9.03	8.49	8.97	0.41
SNR(dB)	4.735	4.676	5.05	3.976	4.39	4.57	0.40



Graph 5.2(a): Insignificant Parameters from table 5.2(a)

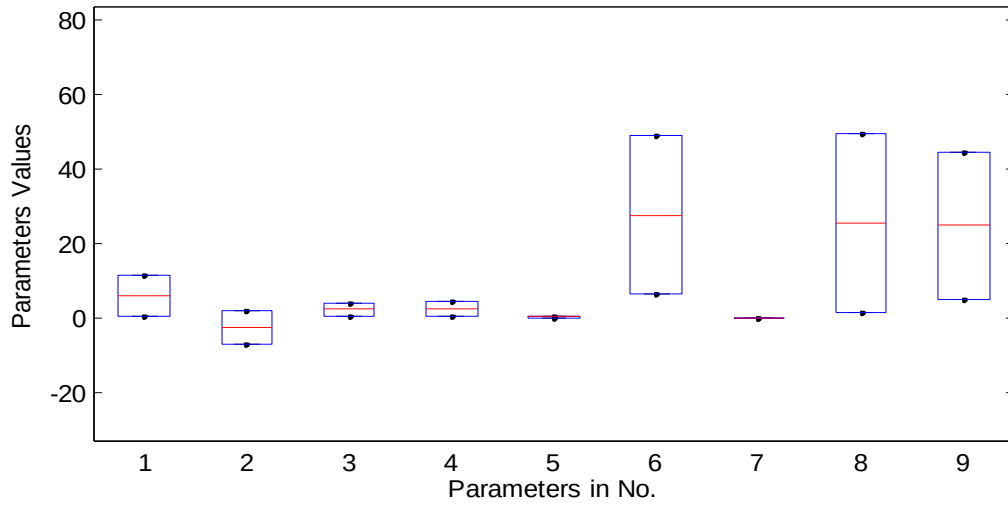


Graph 5.2(b): Significant Parameters from table 5.2(a)

By LPC Spectrum Analysis

Table 5.2(b): Parameters of word 'POLICE'

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	10.9888	12.3019	10.2637	11.293	11.9283	11.36	0.80
MIN	-6.5296	-10.328	-5.3881	-5.3914	-7.549	-7.04	2.05
STD	3.7027	5.3747	4.064	4.2831	3.2934	4.14	0.78
RMS	3.6882	5.3538	4.0481	4.2935	4.3821	4.35	0.62
SLOP	0.42	0.91	0.48	0.55	0.65	0.60	0.19
WIDTH	55	39	54	50	46	48.80	6.53
FRQ	0.0175	0.0195	0.0195	0.0185	0.0183	0.02	0.00
MAXMAG	48.25	51.83	49.27	47.29	49.58	49.24	1.70
A.TION(dB)	42.4	39.1	52.9	44.2	43.3	44.38	5.14



Graph 5.2(c): Significant Parameters from table 5.2(b)

3 BOOM

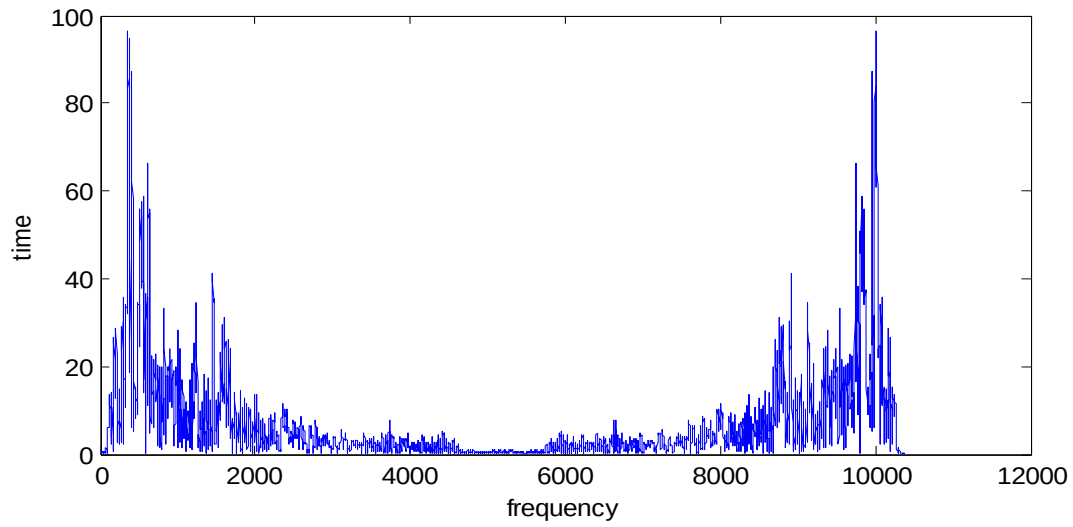
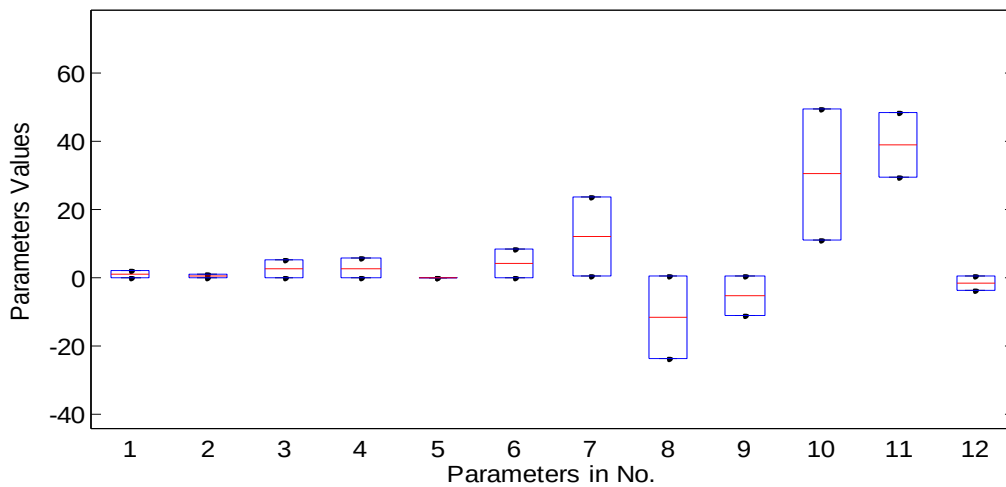


Figure 5.3: Spectrum of 'BOOM' word

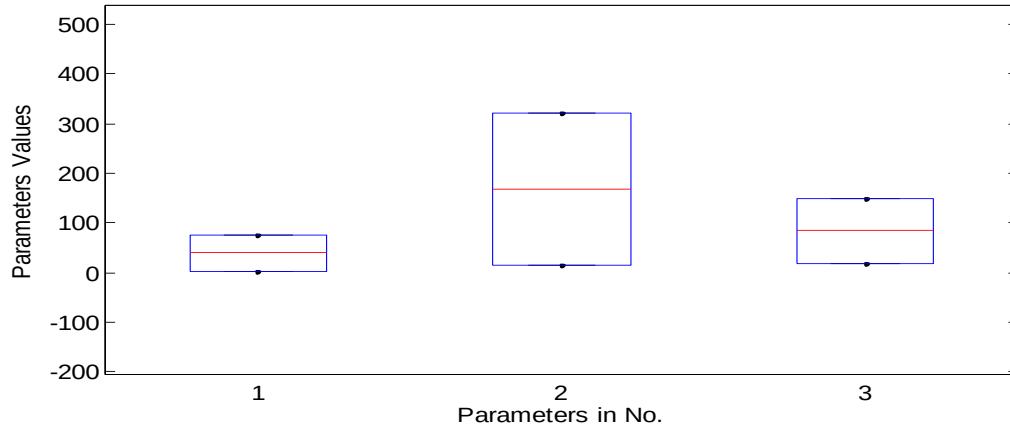
By Spectrum Analysis

Table 5.3(a): Parameters of word 'BOOM'

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	2.075	2.0567	2.0649	2.0592	2.08	2.07	0.01
MEDIAN	1.0177	0.9084	0.998	0.9723	1.0011	0.98	0.04
STD	4.8957	5.6368	5.2845	5.0534	4.9621	5.17	0.30
RMS	5.3171	6.0001	5.5586	5.687	5.4682	5.61	0.26
MAX	73.3352	79.7461	75.689	78.836	75.193	76.56	2.66
MIN	0.0022	0.002	0.0021	0.002	0.0021	0.002080	0.000084
SLOP	8.2	8.4	8.3	8.4	8.3	8.32	0.08
WIDTH	25	23	24	23	24	23.8	0.84
P.FREQ(Hz)	316.41	294.92	333.98	326.17	330.25	320.346	15.65
P. AMP(dB)	-22.77	-23.93	-24.03	-23.41	-23.76	-23.58	0.51
T. POW(dB)	-10.51	-10.3	-11.01	-11.76	-11.37	-10.99	0.60
THD(%)	50.24	30.21	56.71	55.51	55.38	49.61	11.13
THD+N (%)	144.8	165.09	170	130.75	136.04	149.336	17.45
IMD(%)	25.27	30.35	41.16	99.08	45.27	48.226	29.54
SNR(dB)	-3.21	-4.35	-4.27	-3.41	-3.47	-3.742	0.53



Graph 5.3(a): Insignificant Parameters from table 5.3(a)

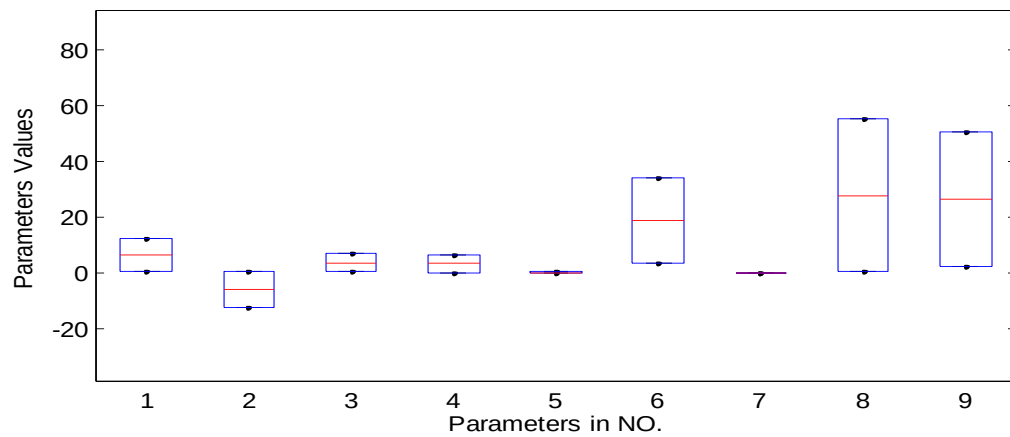


Graph 5.3(b): Significant Parameters from table 5.3(a)

By LPC Spectrum Analysis

Table 5.3(b): Parameters of word ‘BOOM’

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	12.0372	13.2221	12.089	13.011	12.0512	12.48	0.58
MIN	-11.4464	-12.7457	-13.289	-12.351	-12.5953	-12.49	0.68
STD	7.0274	7.7748	6.789	6.391	6.6133	6.92	0.53
RMS	6.9999	6.7288	6.892	6.429	6.5875	6.73	0.23
SLOP	0.69	0.69	0.68	0.69	0.72	0.69	0.02
WIDTH	31	34	32	33	40	34.00	3.54
FRQ	0.0175	0.0175	0.0172	0.0175	0.0136	0.017	0.002
MAXMAG	55.78	55.41	56.28	54.78	54.04	55.26	0.87
A.TION(dB)	53.7	52.8	49.5	48.3	50.4	50.94	2.26



Graph 5.3(c): Significant Parameters from table 5.3(b)

4 BOMB

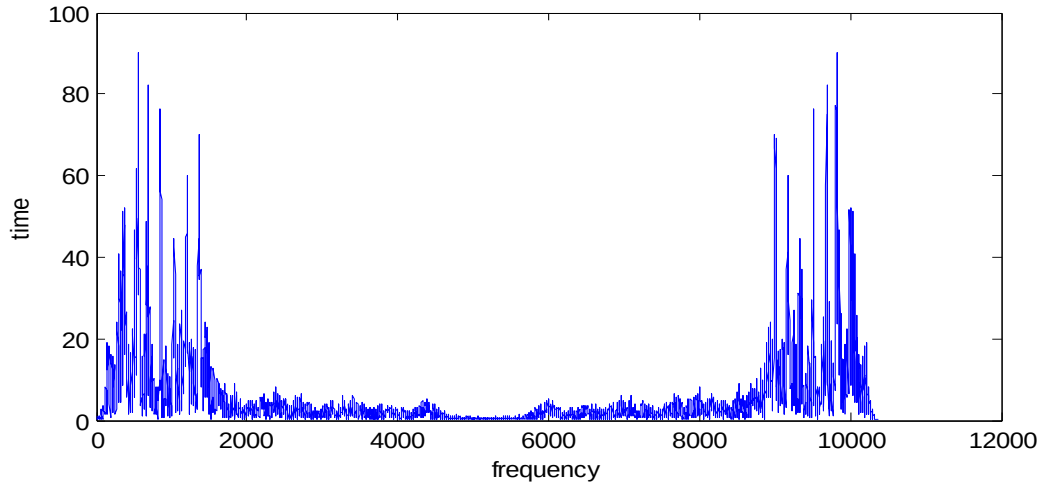
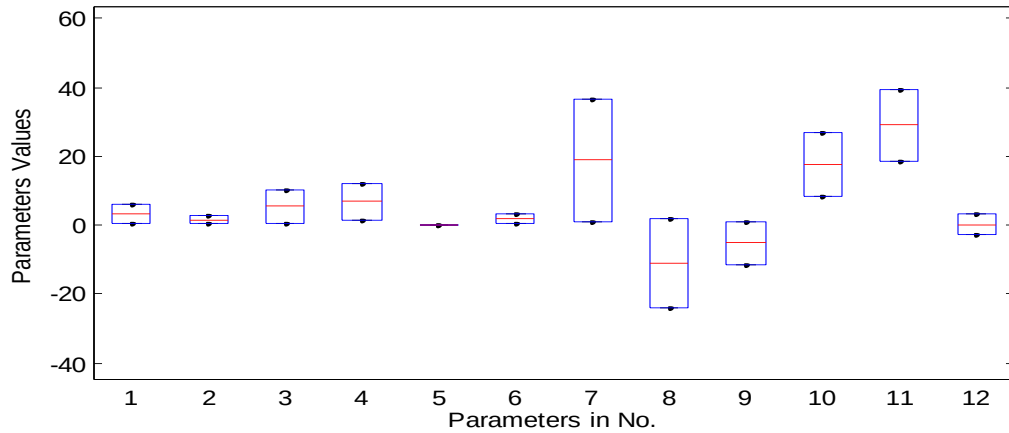


Figure 5.4: Spectrum of 'BOMB' word

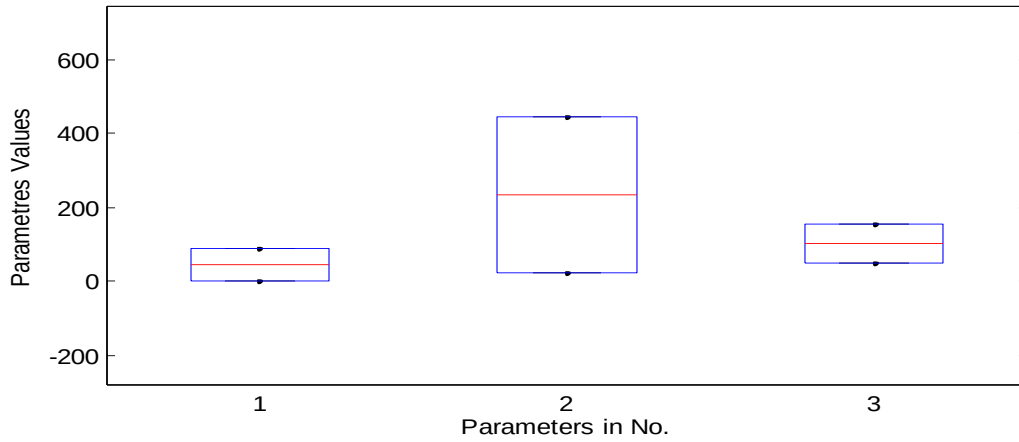
By Spectrum Analysis

Table 5.4(a): Parameters of word 'BOMB'

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	5.896	6.035	6.85	6.22	6.128	6.23	0.37
MEDIAN	2.67	2.51	3.01	2.164	2.73	2.62	0.31
STD	9.078	10.22	10.83	10.9	10.36	10.28	0.73
RMS	10.82	11.87	12.81	12.63	12.45	12.12	1.61
MAX	88.77	90.13	85.67	95.24	91.62	90.29	3.53
MIN	0.012	0.0078	0.014	0.0043	0.014	0.010	0.004
SLOP	3.95	3.06	2.99	3.87	3.62	3.50	0.45
WIDTH	35	38	37	37	36	36.60	1.14
P.FREQ(Hz)	414.06	425.78	458.98	443.56	476.56	443.79	25.06
P.AMP(dB)	-27.18	-22.94	-22.72	-22.42	-24.2	-23.89	1.96
T.POW(dB)	-12.11	-12.18	-10.79	-10.3	-11.09	-11.29	0.83
THD(%)	28.59	20.35	20.3	23.84	40.59	26.73	8.46
THD+N (%)	199.06	82.07	124.56	190.9	181.35	155.59	50.42
IMD(%)	62.89	53.36	27.44	35.53	18.27	39.50	18.38
SNR(dB)	-5.98	1.716	-1.9	-3.39	-5.171	-2.95	3.05



Graph 5.4(a): Insignificant Parameters from table 5.4(a)

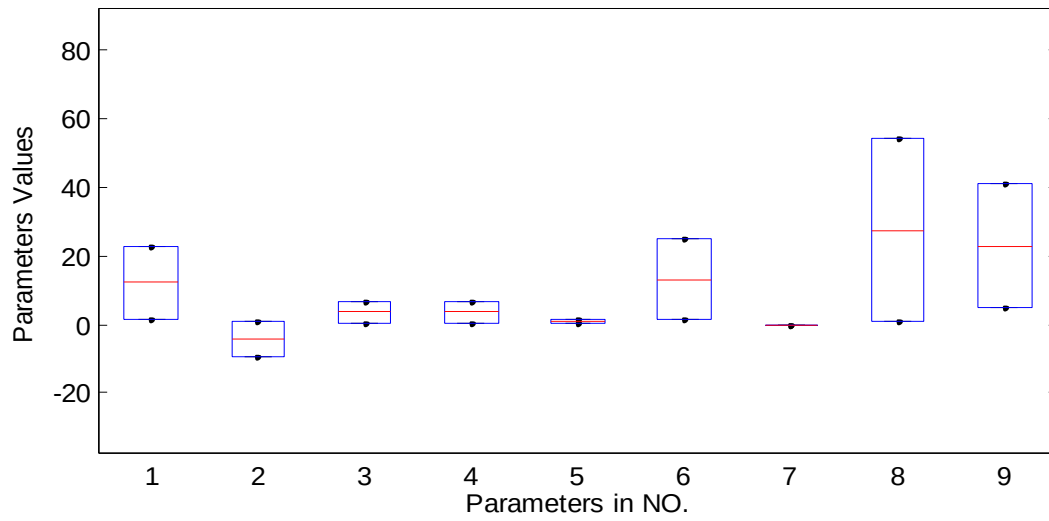


Graph 5.4(b): Significant Parameters from table 5.4(a)

By LPC Spectrum Analysis

Table 5.4(b): Parameters of word 'BOMB'

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	22.8333	21.2102	24.1983	25.1019	21.5065	22.97	1.68
MIN	-8.6082	-10.24	-8.9592	-10.687	-9.4975	-9.60	0.87
STD	6.807	6.7743	6.8337	7.0977	6.1289	6.73	0.36
RMS	6.7804	6.7478	6.807	7.0702	6.105	6.70	0.36
SLOP	1.76	1.29	2.04	1.25	1.43	1.55	0.34
WIDTH	26	24	26	23	26	25.00	1.41
FRQ at	0.0117	0.0156	0.0156	0.0097	0.0117	0.013	0.003
MAXMAG	55.17	54.53	53.07	55.01	53.06	54.17	1.03
A.TION(dB)	37.4	45.6	46.3	36.8	38.5	40.92	4.64



Graph 5.4(c): Significant Parameters from table 5.4(b)

5 BUS

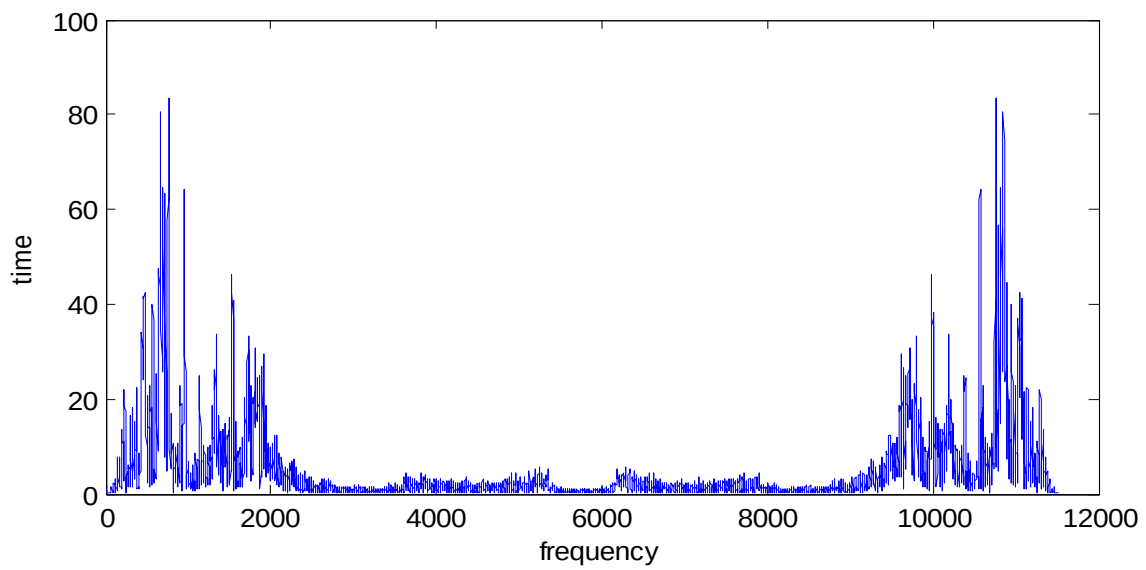
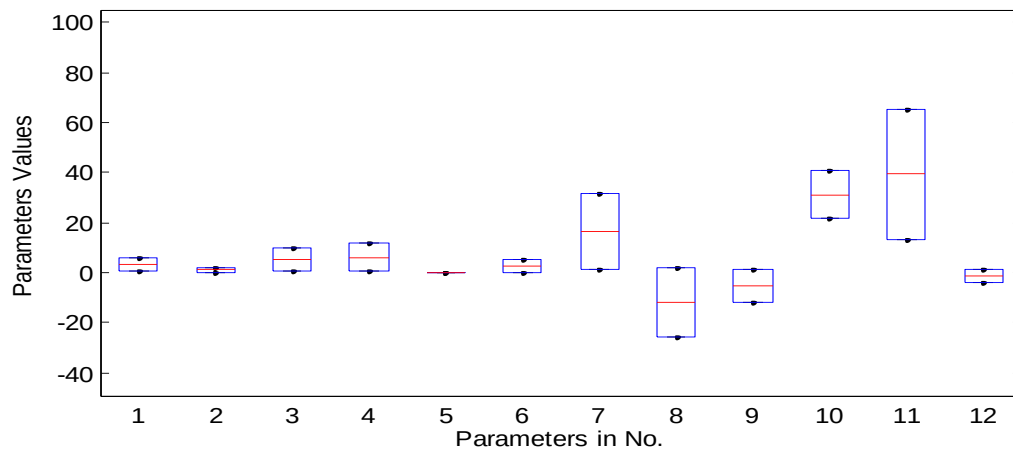


Figure 5.5: Spectrum of 'BUS' word

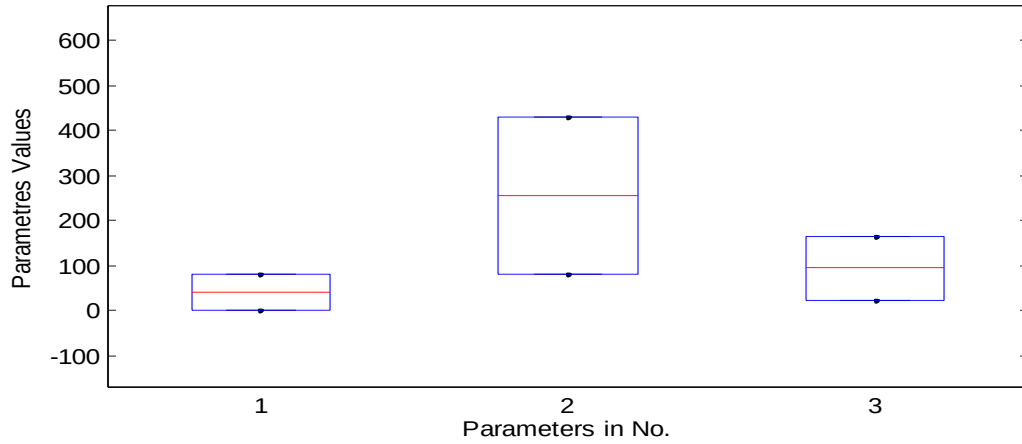
By Spectrum Analysis

Table 5.5(a): Parameters of word 'BUS'

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	5.58	6.63	6.22	6.34	6.09	6.17	0.39
MEDIAN	1.84	2.45	1.99	2.01	2.05	2.07	0.23
STD	9.65	10.41	10.65	10.35	10.01	10.21	0.39
RMS	11.15	12.34	12.32	12.33	11.99	12.03	0.51
MAX	83.3	78.97	78.88	79.22	80.29	80.13	1.86
MIN	0.013	0.011	0.015	0.014	0.012	0.013	0.002
SLOP	5.53	5.3	5.54	5.5	5.2	5.41	0.15
WIDTH	31	33	31	32	30	31.40	1.14
P.FREQ(Hz)	462.89	511.72	484.38	332.03	349.61	428.13	81.79
P. AMP(dB)	-27.26	-26.65	-25.77	-25.01	-22.41	-25.42	1.89
T. POW(dB)	-13.48	-12.92	-12.74	-10.52	-9.78	-11.89	1.63
THD(%)	27.57	27.17	29.17	77.43	43.8	41.03	21.49
THD+N (%)	183.38	163.34	161.25	185.09	129.62	164.54	22.41
IMD(%)	56.14	63.07	59.56	88.42	60.33	65.50	13.05
SNR(dB)	-4.62	-4.26	-4.151	-5.3	-2.25	-4.12	1.14



Graph 5.5(a): Insignificant Parameters from table5.5(a)

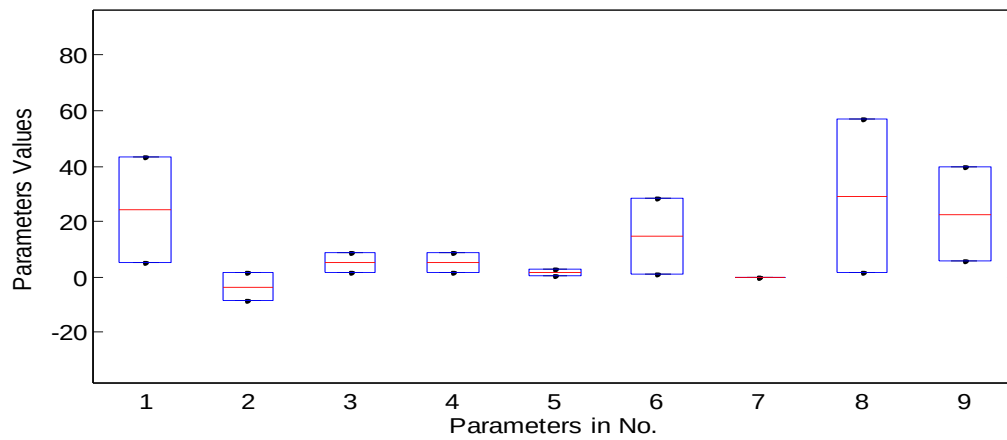


Graph 5.5(b): Significant Parameters from table 5.5(a)

By LPC Spectrum Analysis

Table 5.5(b): Parameters of word 'BUS'

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	48.3242	49.0291	41.5901	38.3612	39.3152	43.32	5.03
MIN	-10.056	-9.6018	-8.2148	-7.0803	-7.1302	-8.42	1.38
STD	10.1682	10.2215	8.938	7.0491	7.61	8.80	1.45
RMS	10.129	10.1822	8.9034	7.0294	7.5803	8.76	1.44
SLOP	3.16	2.62	2.64	2.89	2.25	2.71	0.34
WIDTH	30	28	28	27	28	28.20	1.10
FRQ at	0.0097	0.0117	0.0117	0.0136	0.0117	0.012	0.001
MAXMAG	58.56	58.47	57.38	54.43	55.73	56.91	1.80
A.TION(dB)	35.6	34.8	36.4	44.5	46.8	39.62	5.59



Graph 5.5(c): Significant Parameters from table 5.5(b)

6 BUT

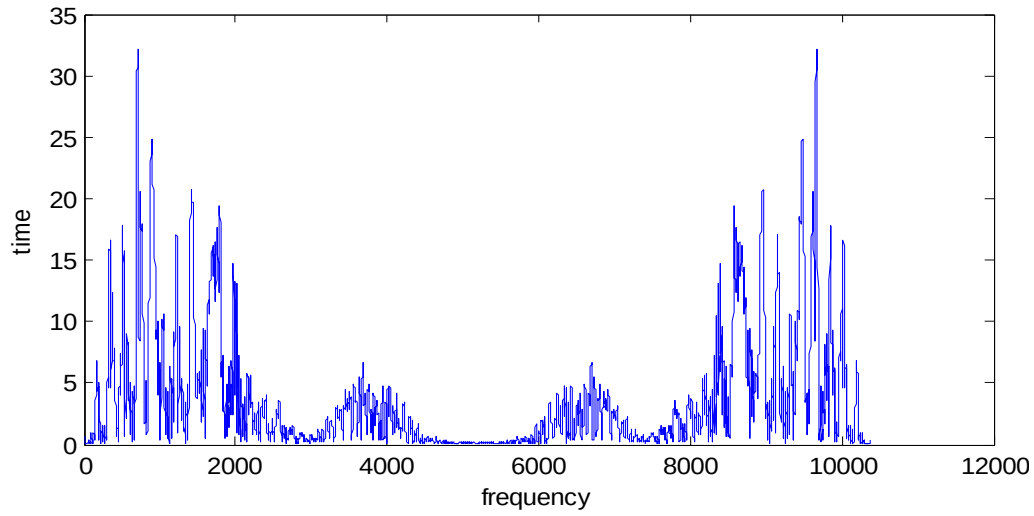
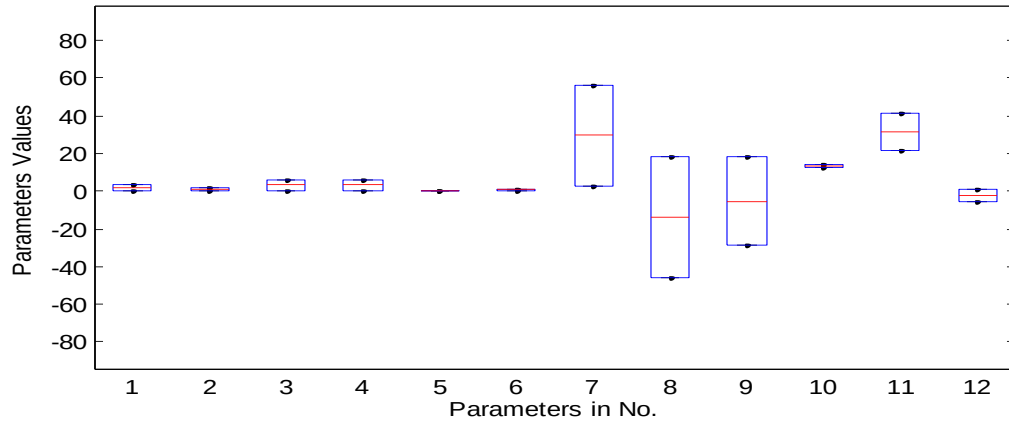


Figure 5.6: Spectrum of 'BUT' word

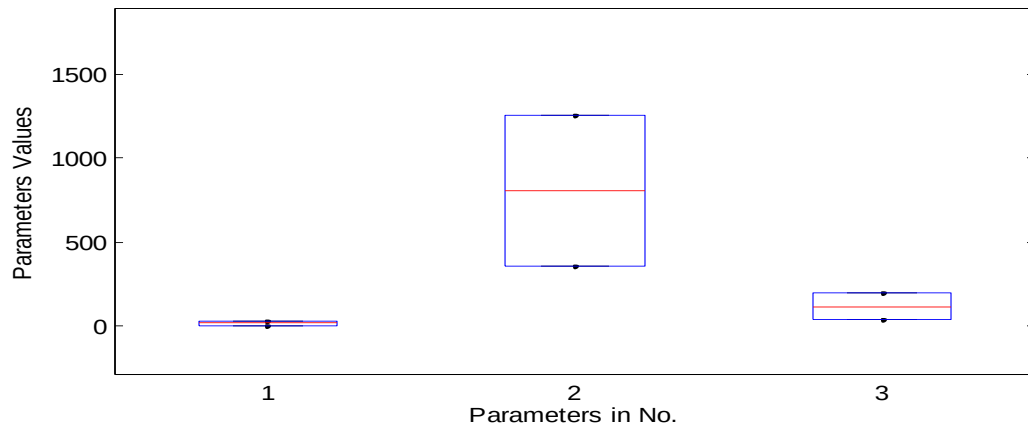
By Spectrum Analysis

Table 5.6(a): Parameters of word 'BUT'

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	3.235	3.766	3.824	3.601	3.681	3.62	0.23
MEDIAN	1.063	1.959	1.518	1.492	1.873	1.58	0.36
STD	6.527	5.501	5.051	5.302	5.823	5.64	0.57
RMS	6.285	6.217	6.335	6.301	6.298	6.29	0.04
MAX	35.15	32.106	30.78	31.39	33.35	32.56	1.74
MIN	0.0029	0.0052	0.0015	0.0035	0.003	0.003	0.001
SLOP	1.85	0.93	0.45	0.89	1.01	1.03	0.51
WIDTH	56	52	60	58	56	56.40	2.97
P.FREQ(Hz)	611.33	1382.81	1470.7	1394.53	1399.08	1251.69	359.63
P. AMP(dB)	-78.12	-40.39	-38.81	-35.76	-37.88	-46.19	17.93
T. POW(dB)	-62.05	-21.22	-22.37	-18.68	-20.59	-28.98	18.53
THD(%)	36.12	15.41	7.82	5.82	6.87	14.41	12.71
THD+N (%)	175.15	258.48	180.17	191.18	185.49	198.09	34.28
IMD(%)	73.39	13.54	34.99	44.32	38.78	41.00	21.54
SNR(dB)	-4.86	-8.2	-5.11	-5.62	-5.39	-5.84	1.35



Graph 5.6(a): Insignificant Parameters from table 5.6(a)

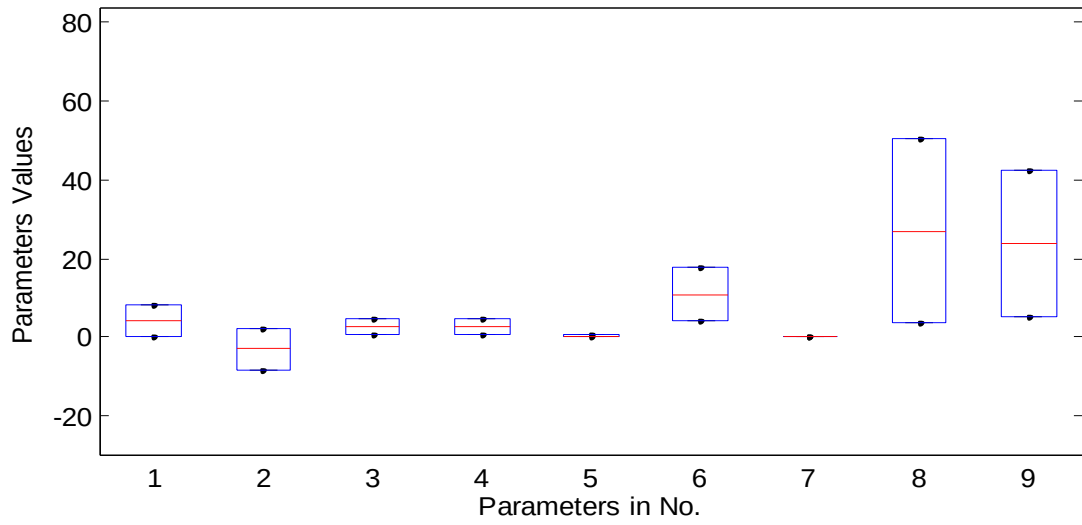


Graph 5.6(b): Significant Parameters from table 5.6(a)

By LPC Spectrum Analysis

Table 5.6(b): Parameters of word 'BUT'

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	8.2899	8.389	8.8381	8.9074	7.8665	8.46	0.43
MIN	-7.9974	-8.19	-9.6153	-10.748	-4.871	-8.28	2.21
STD	4.0329	4.9529	5.2137	5.6943	4.87	4.95	0.61
RMS	4.0173	4.8926	5.1934	5.6722	4.8588	4.93	0.60
SLOP	0.71	0.7	0.66	0.706	0.706	0.70	0.02
WIDTH	24	16	16	14	20	18.00	4.00
FRQ	0.0097	0.0136	0.0136	0.0117	0.0136	0.012	0.002
MAXMAG	47.36	52.68	52.75	53.84	45.12	50.35	3.86
A.TION(dB)	35.5	38.8	42.3	46.2	48.3	42.22	5.24



Graph 5.6(c): Significant Parameters from table 5.6(b)

7 PUT

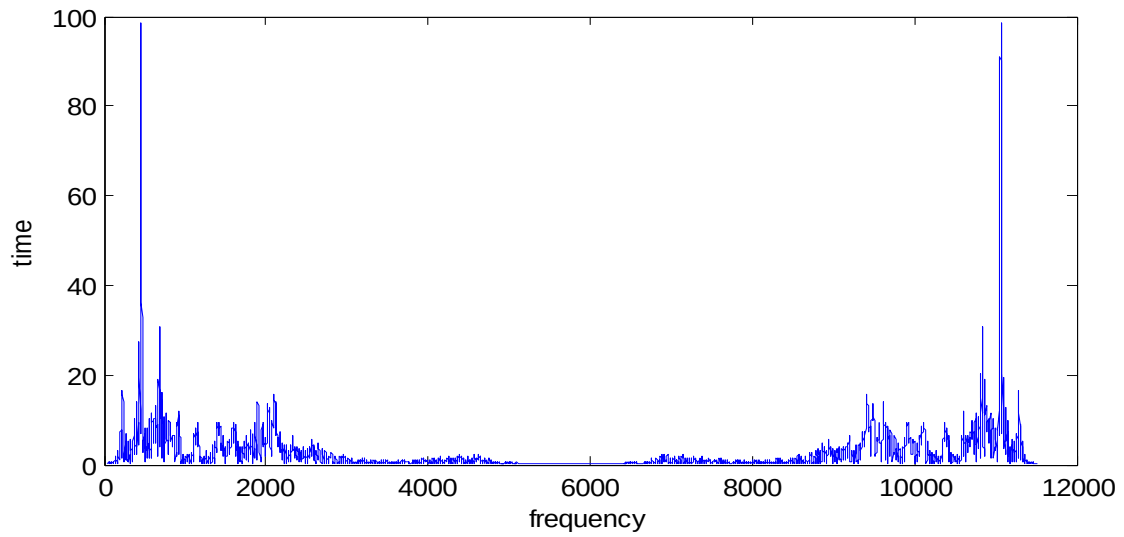
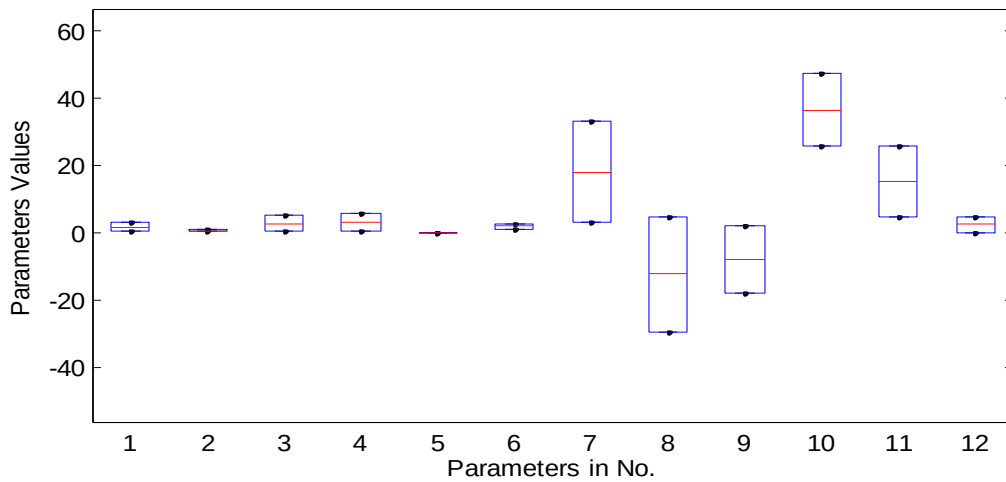


Figure 5.7: Spectrum of 'PUT' word

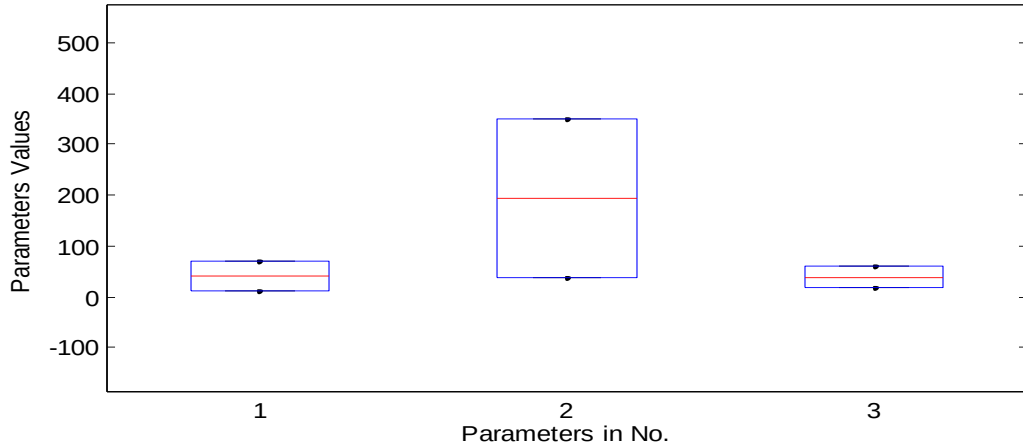
By Spectrum Analysis

Table 5.7(a): Parameters of word ‘PUT’

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	3.705	2.926	2.481	2.67	3.027	2.96	0.47
MEDIAN	1.178	1.396	0.85	0.99	1.179	1.12	0.21
STD	5.104	4.684	4.741	5.362	5.399	5.06	0.34
RMS	6.306	5.523	5.351	5.991	6.201	5.87	0.42
MAX	56.91	62.44	80.73	84.77	68.46	70.66	11.85
MIN	0.0046	0.0081	0.0019	0.0023	0.0039	0.004	0.002
SLOP	1.038	2.59	2.92	3.56	2.9	2.60	0.94
WIDTH	35	30	36	30	34	33.00	2.83
P.FREQ(Hz)	344.34	410.7	316.41	316.41	359.38	349.45	38.92
P. AMP(dB)	-28.29	-37.03	-28.8	-23.74	-30.62	-29.70	4.82
T.POW(dB)	-18.47	-20.2	-20.01	-16.64	-16.24	-18.31	1.84
THD(%)	38.13	81.67	21.007	15..50	49.87	47.67	25.58
THD+N (%)	39.48	60.51	84.31	47.4	72.02	60.74	18.12
IMD(%)	25.74	18.54	28.21	24.23	31.64	25.67	4.87
SNR(dB)	1.87	-5.21	1.48	6.48	-4.71	-0.02	4.92



Graph 5.7(a): Insignificant Parameters from table 5.7(a)

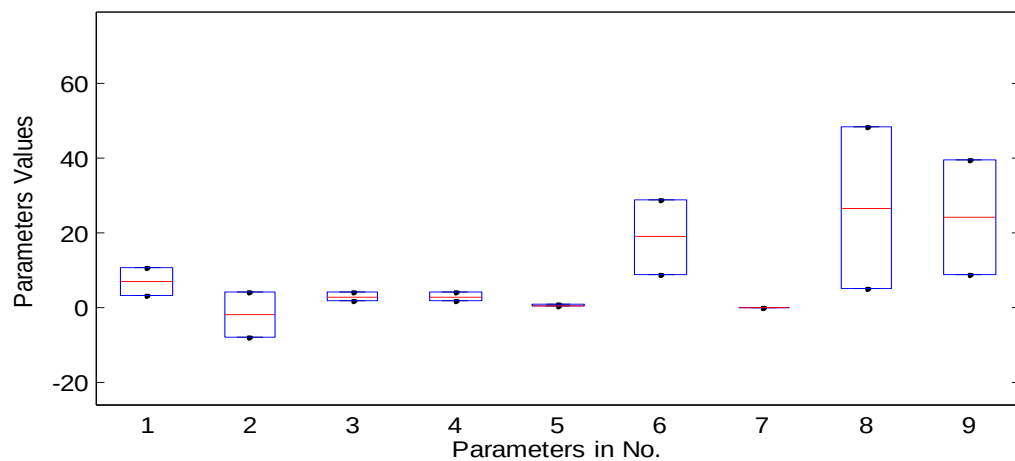


Graph 5.7(b): Significant Parameters from table 5.7(a)

By LPC Spectrum Analysis

Table 5.7(b): Parameters of word 'PUT'

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	9.467	10.2249	12.8964	12.6386	5.9805	10.24	2.81
MIN	-8.429	-7.5547	-8.1398	-12.96	-3.3854	-8.09	3.40
STD	4.792	4.4424	3.8736	6.104	2.1324	4.27	1.45
RMS	4.582	4.425	3.8589	6.0805	2.124	4.21	1.43
SLOP	0.86	0.71	1.05	0.81	0.28	0.74	0.29
WIDTH	26	25	24	25	42	28.40	7.64
FRQ	0.0098	0.0156	0.042	0.0097	0.0191	0.019	0.013
MAXMAG	49.56	49.7	47.22	54.5	42.11	48.62	4.50
A.TION(dB)	40.2	48.4	30	35	45.2	39.76	7.46



Graph 5.7(c): Significant Parameters from table 5.7(b)

8 FILL

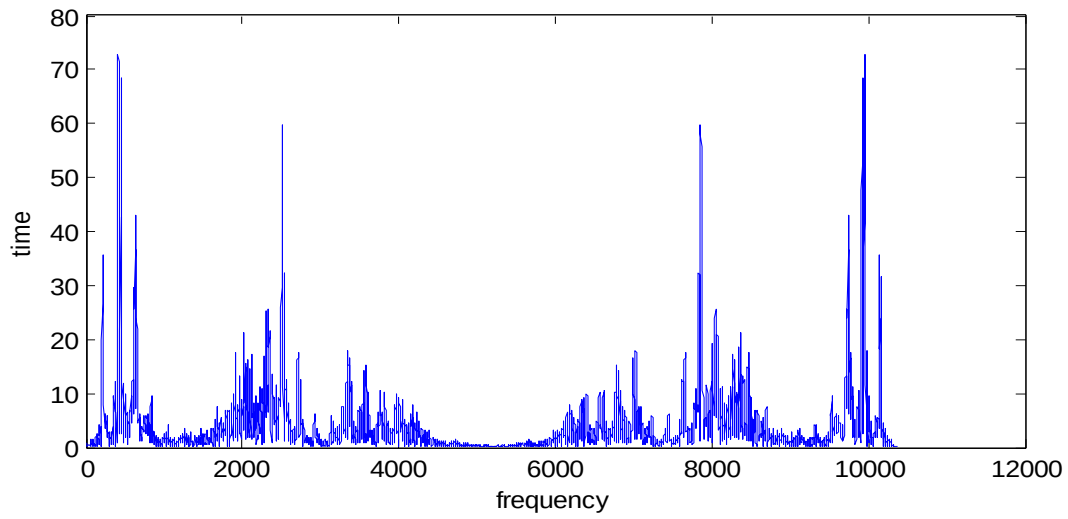
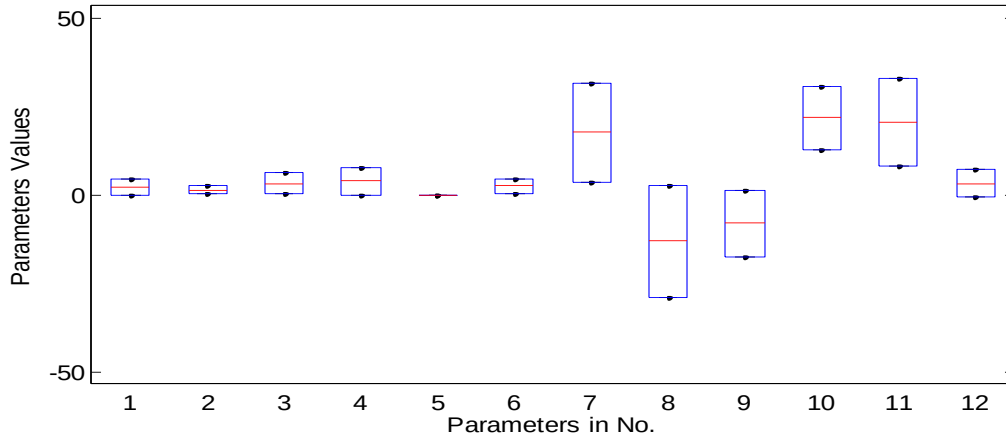


Figure 5.8: Spectrum of 'FILL' word

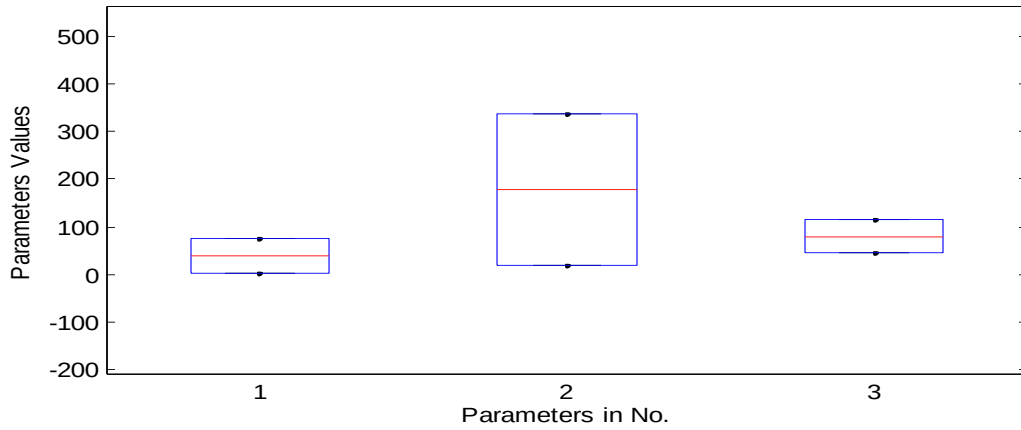
By Spectrum Analysis

Table 5.8(a): Parameters of word 'FILL'

	S1	S2	S3	S4	Avg	Stdev
MEAN	4.749	4.311	4.54	4.431	4.51	0.19
MEDIAN	2.822	2.305	2.748	2.695	2.64	0.23
STD	6.085	6.775	6.612	6.49	6.49	0.29
RMS	7.719	8.03	8.021	7.98	7.94	0.15
MAX	69.76	72.83	78.92	75.46	74.24	3.89
MIN	0.0056	0.0015	0.0052	0.0047	0.004	0.002
SLOP	4.36	4.23	5.05	5	4.66	0.43
WIDTH	37	31	29	30	31.75	3.59
P. FREQ(Hz)	359.38	345.7	328.12	312.5	336.43	20.45
P. AMP(dB)	-30.62	-29.18	-30.61	-24.52	-28.73	2.89
T. POW(dB)	-16.19	-18.46	-18	-16.14	-17.20	1.21
THD(%)	22.63	26.65	26.69	20.5	24.12	3.07
THD+N (%)	115.72	105.57	118.69	91.96	107.99	12.07
IMD(%)	31.56	34.01	32.29	43.62	35.37	5.60
SNR(dB)	-4.69	-0.345	-1.408	-2.157	-2.15	3.70



Graph 5.8(a): Insignificant Parameters from table 5.8(a)

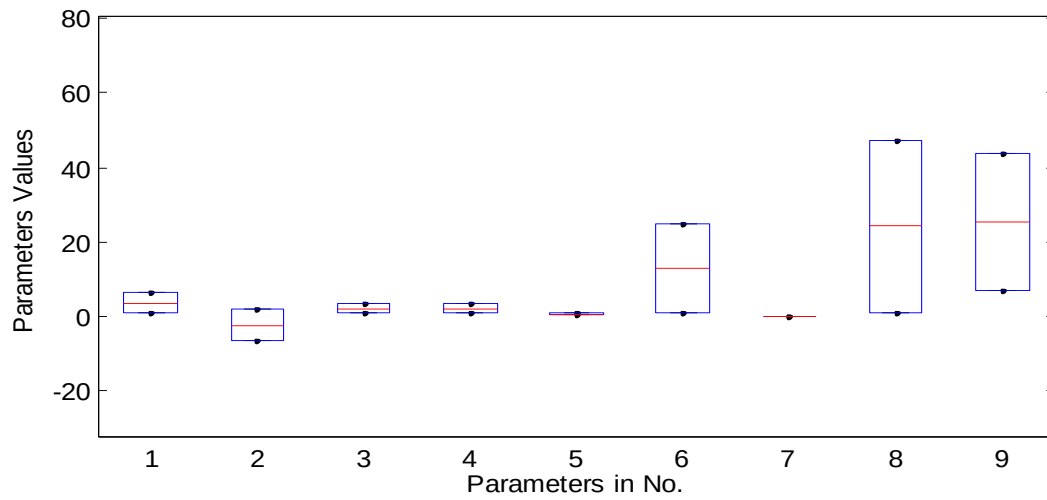


Graph 5.8(b): Significant Parameters from table 5.8(a)

By LPC Spectrum Analysis

Table 5.9(b): Parameters of word 'FILL'

	S1	S2	S3	S4	Avg	Stdev
MAX	5.5731	6.3526	7.8783	6.078	6.47	0.99
MIN	-3.9771	-6.0073	-8.229	-7.6212	-6.46	1.90
STD	2.7445	3.3116	4.5701	3.5789	3.55	0.76
RMS	2.7337	3.2989	4.5522	3.5649	3.54	0.76
SLOP	0.59	0.55	0.53	0.43	0.53	0.07
WIDTH	26	25	24	24	24.75	0.96
FRQ	0.0136	0.0097	0.021	0.0136	0.014	0.005
MAXMAG	45.67	47.28	48.4	47.06	47.10	1.12
A.TION(dB)	49.1	33.1	46.5	45.7	43.60	7.15



Graph 5.8(c): Significant Parameters from table 5.8(b)

9 KILL

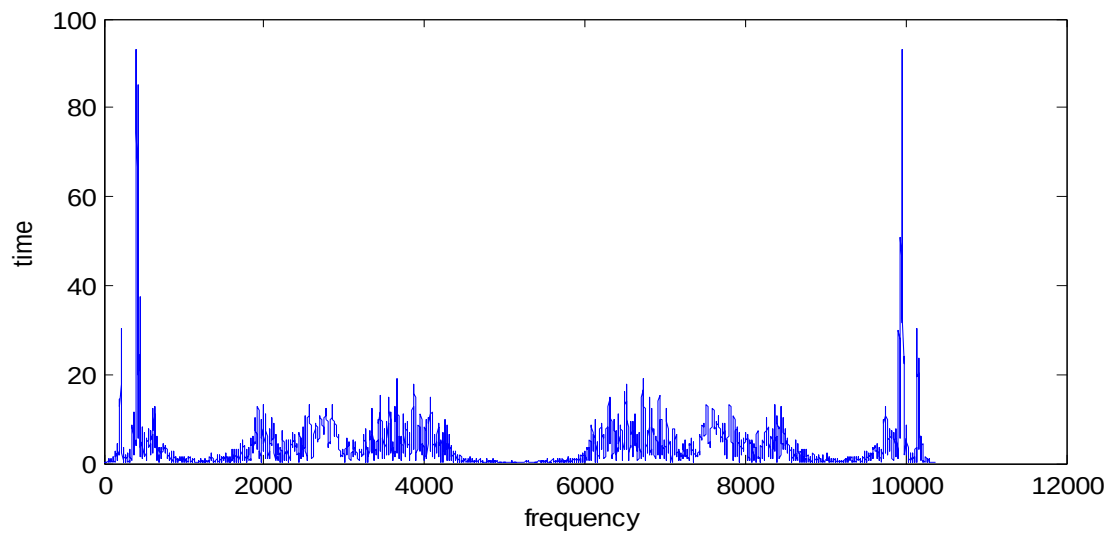
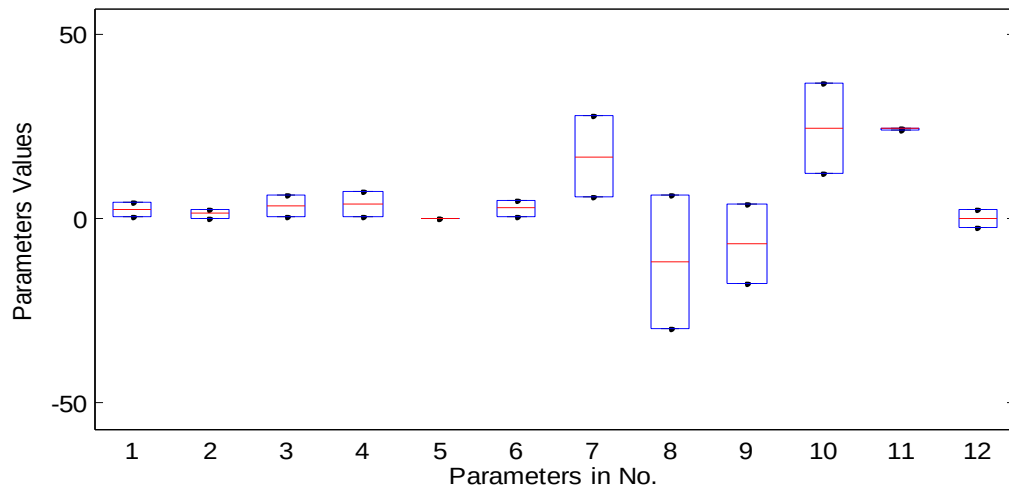


Figure 5.9: Spectrum of 'KILL' word

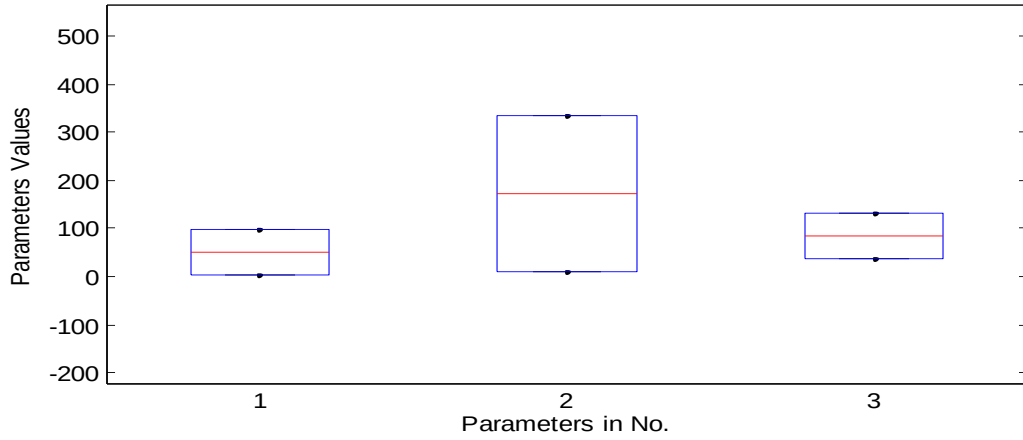
By Spectrum Analysis

Table 5.9(a): Parameters of word ‘KILL’

	S1	S2	S3	S4	Avg	Stdev
MEAN	5.24	3.85	4.17	4.84	4.53	0.63
MEDIAN	3.13	2.395	2.572	2.834	2.73	0.32
STD	7.693	5.67	6.45	6.29	6.53	0.85
RMS	8.313	6.86	7.68	6.97	7.46	0.68
MAX	100.99	92.89	98.05	94.72	96.66	3.59
MIN	0.0152	0.0076	0.0057	0.006	0.009	0.004
SLOP	5.4	4.44	5.6	5.57	5.25	0.55
WIDTH	24	36	23	29	28.00	5.94
P. FREQ(Hz)	318.36	339.84	343.75	333.98	333.98	11.16
P. AMP(dB)	-20.19	-32.35	-32.77	-33.5	-29.70	6.36
T. POW(dB)	-11.67	-20.08	-19.48	-19.05	-17.57	3.96
THD(%)	34.28	38.23	33.16	32.2	34.47	2.65
THD+N (%)	130.96	134.89	136.16	140.01	135.51	3.73
IMD(%)	16.82	15.66	12.46	17.2	15.54	2.15
SNR(dB)	-2.622	-2.6	-2.68	-4.61	-3.13	0.99



Graph 5.9(a): Insignificant Parameters from table 5.9(a)

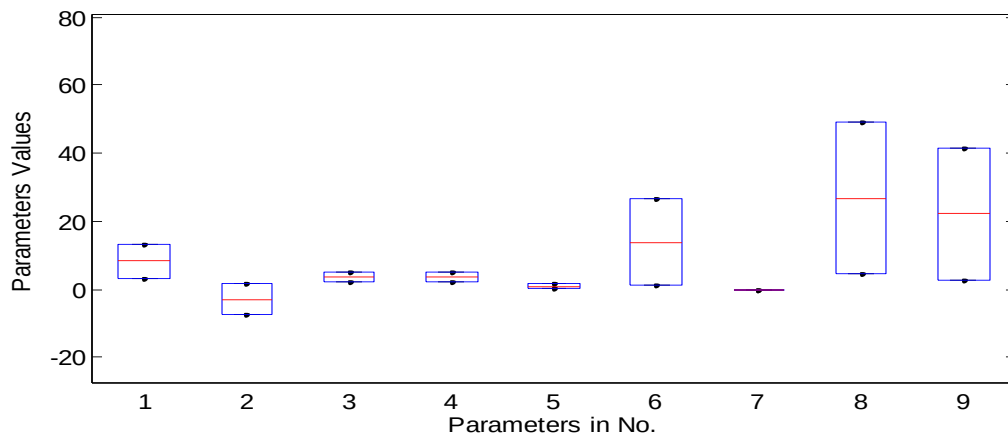


Graph 5.9(b): Significant Parameters from table 5.9(a)

By LPC Spectrum Analysis

Table 5.9(b): Parameters of word ‘KILL’

	S1	S2	S3	S4	Avg	Stdev
MAX	12.3251	16.1165	15.6312	8.8867	13.24	3.36
MIN	-7.7698	-9.1522	-8.0065	-5.4609	-7.60	1.55
STD	4.6178	8.295	3.9761	3.3393	5.06	2.22
RMS	4.5998	8.2626	3.9605	3.3262	5.04	2.21
SLOP	1.4	1.55	1.75	1.53	1.56	0.14
WIDTH	26	28	25	27	26.50	1.29
FRQ	0.023	0.0151	0.021	0.021	0.020	0.003
MAXMAG	49.06	55.81	45.54	45.99	49.10	4.74
A.TION(dB)	38.7	43	45	39.9	41.65	2.88



Graph 5.9(c): Significant Parameters from table 5.9(b)

10 SINGH

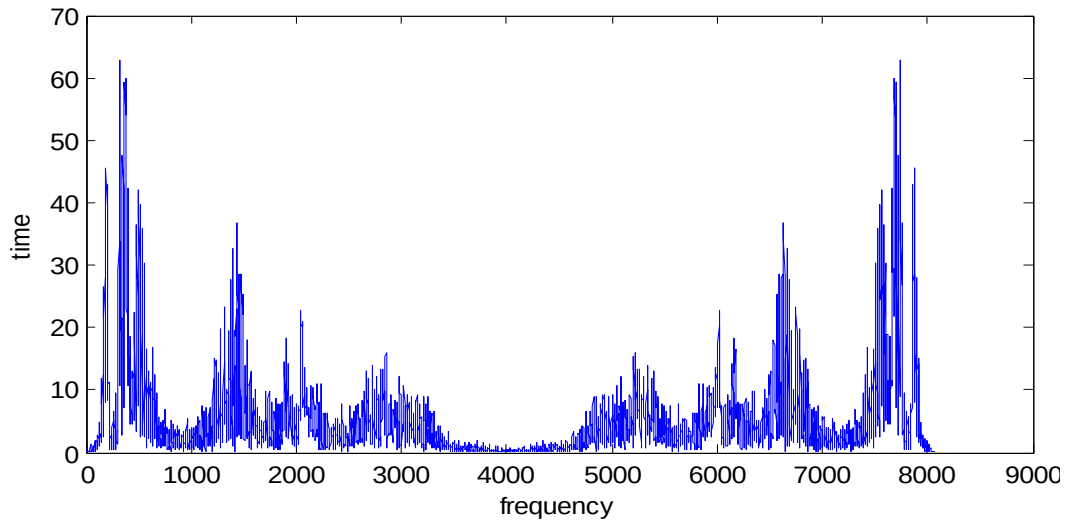
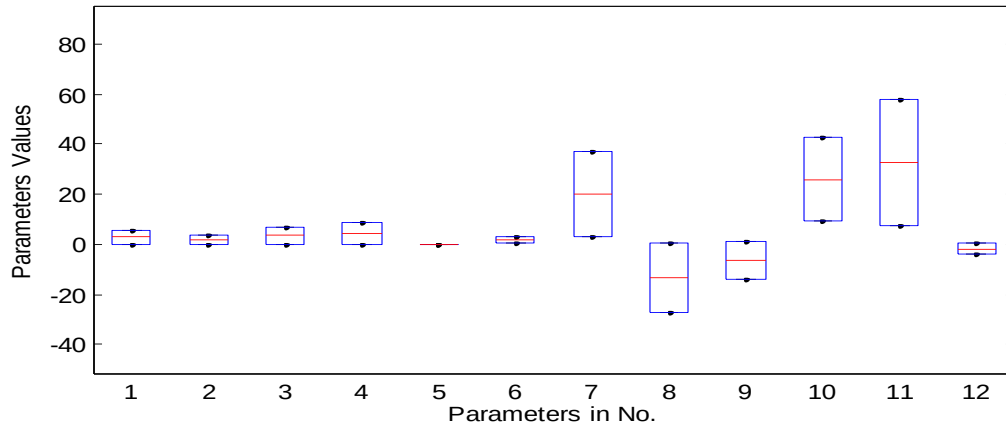


Figure 5.10: Spectrum of 'SINGH' word

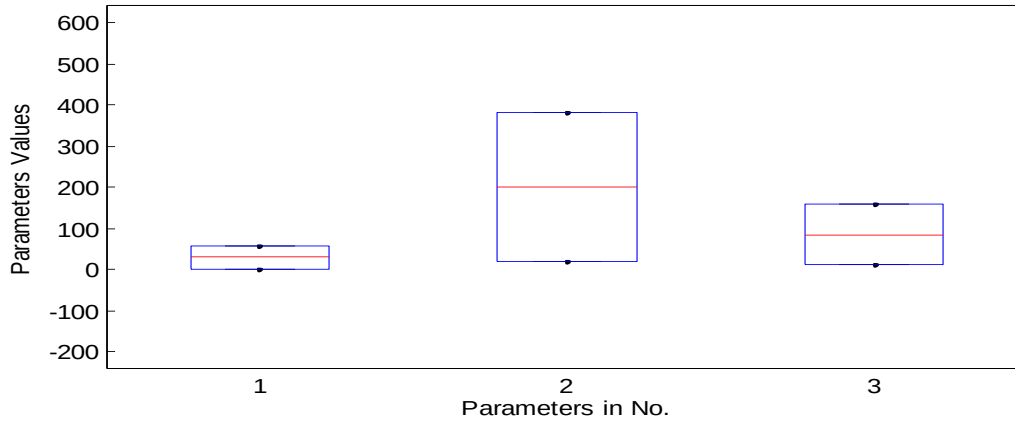
By Spectrum Analysis

Table 5.10(a): Parameters of word 'SINGH'

	S1	S2	S3	S4	Avg	Stdev
MEAN	5.65	5.72	5.31	5.6	5.57	0.18
MEDIAN	3.73	3.64	3.48	3.53	3.60	0.11
STD	6.42	6.68	6.69	6.64	6.61	0.13
RMS	8.55	8.74	8.54	8.63	8.62	0.09
MAX	52.97	55.79	58.88	57.36	56.25	2.52
MIN	0.05	0.0071	0.0077	0.0066	0.02	0.02
SLOP	2.035	3.021	4.001	3.04	3.02	0.80
WIDTH	40	36	33	39	37.00	3.16
P.FREQ(Hz)	404.3	388.67	355.47	373.05	380.37	20.94
P.AMP(dB)	-28.17	-26.51	-28.15	-27.41	-27.56	0.78
T.POW(dB)	-13.44	-13.34	-15.34	-13.19	-13.83	1.01
THD(%)	36.84	44.81	33.96	54.3	42.48	9.12
THD+N (%)	148.51	157.99	157.88	174.16	159.64	10.65
IMD(%)	50.72	52.94	61.83	66	57.87	7.24
SNR(dB)	-3.43	-3.97	-3.967	-4.819	-4.05	0.57



Graph 5.10(a): Insignificant Parameters from table 5.10(a)

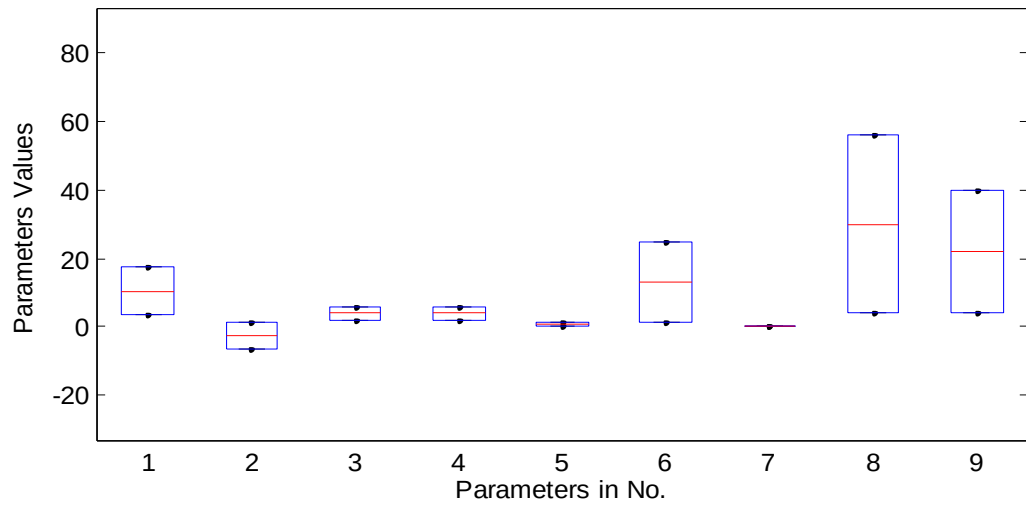


Graph 5.10(b): Significant Parameters from table 5.10(a)

By LPC Spectrum Analysis

Table 5.10(a): Parameters of word ‘SINGH’

	S1	S2	S3	S4	Avg	Stdev
MAX	14.1406	16.7775	15.667	22.3416	17.23	3.57
MIN	-5.1398	-6.7173	-5.783	-8.656	-6.57	1.53
STD	3.822	6.7212	5.6821	8.1101	6.08	1.81
RMS	3.8071	6.6952	5.6639	8.0784	6.06	1.80
SLOP	1.36	1.08	1.25	1.13	1.21	0.13
WIDTH	25	23	24	26	24.50	1.29
FRQ	0.029	0.0117	0.0102	0.0171	0.017	0.009
MAXMAG	46.52	54.82	54.28	55.78	55.78	4.27
A.TION(dB)	35.6	44.1	42.1	37.6	39.85	3.93



Graph 5.10(c): Significant Parameters from table 5.10(b)

11 KING

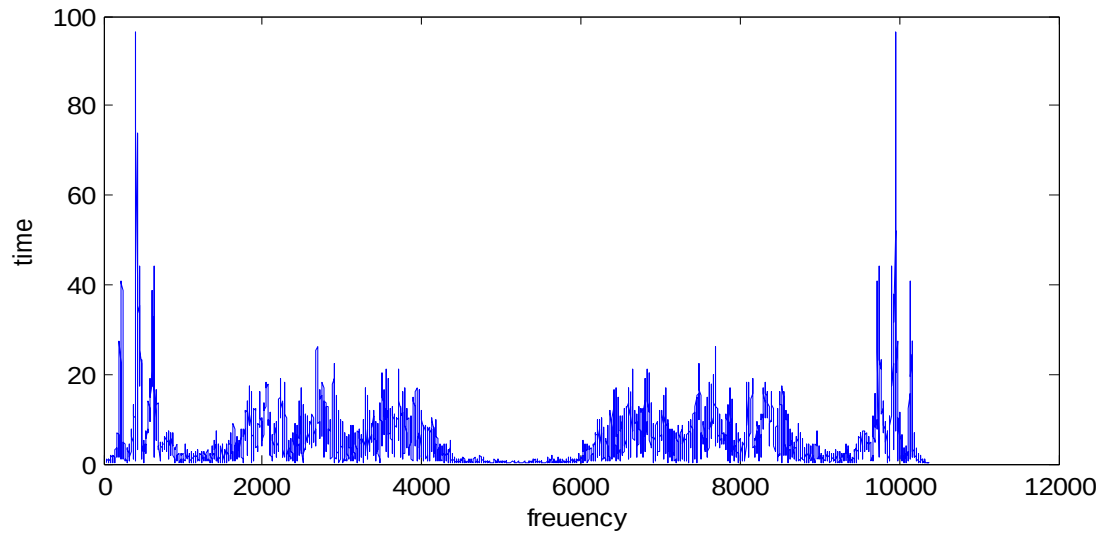
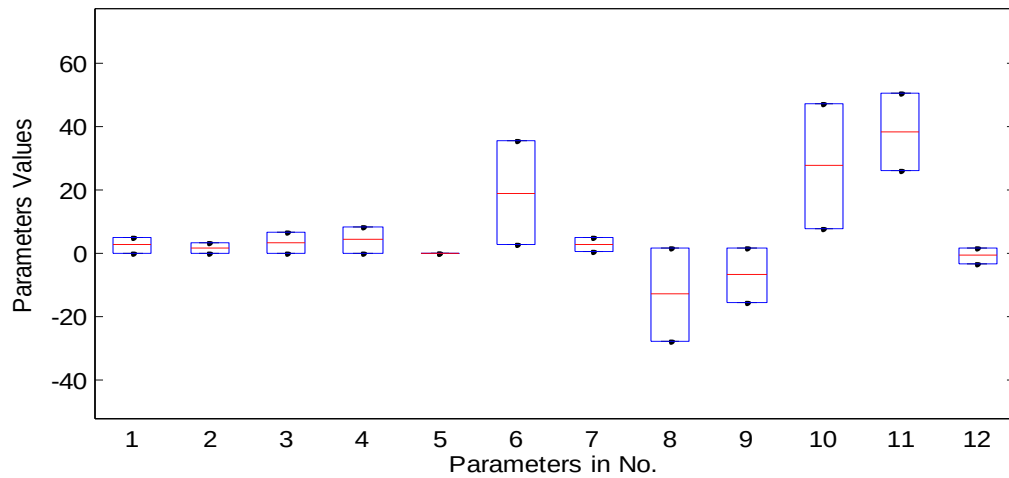


Figure 5.11: Spectrum of 'KING' word

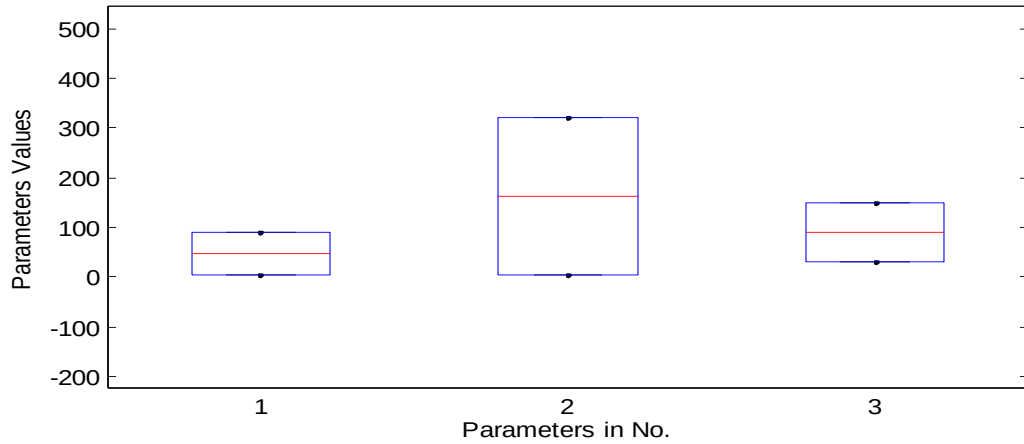
By Spectrum Analysis

Table 5.11(a): Parameters of word 'KING'

	S1	S2	S3	S4	Avg	Stdev
MEAN	5.07	5.56	5.076	5.38	5.27	0.24
MEDIAN	3.35	3.71	3.359	3.4	3.45	0.17
STD	6.44	6.79	6.44	6.5	6.54	0.17
RMS	8.2	8.78	8.206	8.49	8.42	0.28
MAX	87.76	96.49	87.76	88.39	90.10	4.27
MIN	0.0097	0.0068	0.0097	0.0089	0.009	0.001
SLOP	36	32	38	37	35.75	2.63
WIDTH	5	6	5	5	5.25	0.50
P. FREQ(Hz)	326.17	314.45	322.27	316.41	319.83	5.38
P. AMP(dB)	-29.2	-26.5	-26.18	-29.92	-27.95	1.89
T. POW(dB)	-14.36	-14.27	-14.53	-18.06	-15.31	1.84
THD(%)	52.26	43.34	55.76	38.98	47.59	7.76
THD+N (%)	192.02	129.91	151.57	119.61	148.28	32.06
IMD(%)	42	39.09	32.24	89.15	50.62	26.01
SNR(dB)	-5.66	-2.29	-3.61	-1.556	-3.28	1.80



Graph 5.11(a): Insignificant Parameters from table 5.11(a)

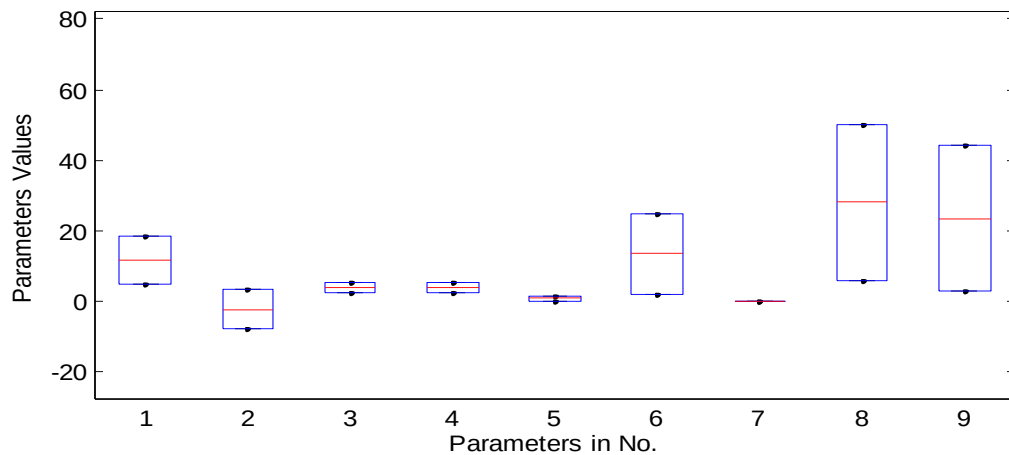


Graph 5.11(b): Significant Parameters from table 5.11(a)

By LPC Spectrum Analysis

Table 5.11(b): Parameters of word 'KING'

	S1	S2	S3	S4	Avg	Stdev
MAX	20.3233	18.166	24.0419	12.1426	18.67	4.98
MIN	-3.6197	-11.2803	-10.1653	-6.344	-7.85	3.53
STD	3.9614	7.4508	7.8328	3.1425	5.60	2.39
RMS	3.9459	7.4218	7.8022	3.1302	5.58	2.38
SLOP	1.73	1.54	1.39	1.54	1.55	0.14
WIDTH	27	22	26	24	24.75	2.22
FRQ	0.032	0.0117	0.0117	0.0175	0.018	0.010
MAXMAG	45.75	55.24	55.23	44.69	50.23	5.80
A.TION(dB)	46.8	46.7	41.1	42.5	44.28	2.91



Graph 5.11(c): Significant Parameters from table 5.11(b)

12 KITE

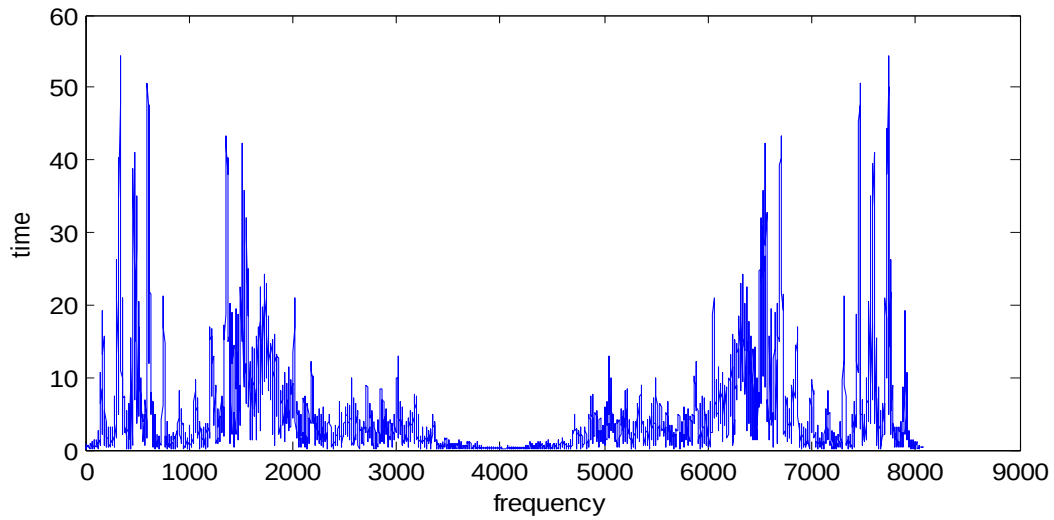
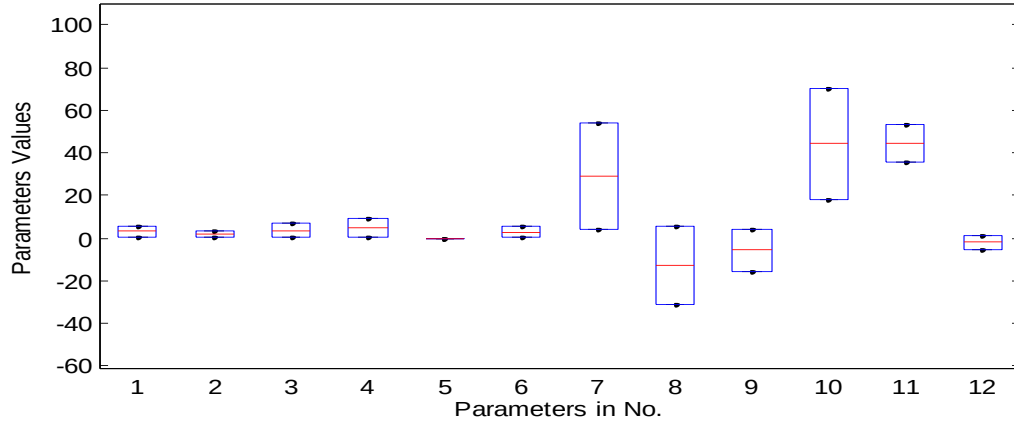


Figure 5.12: Spectrum of 'KITE' word

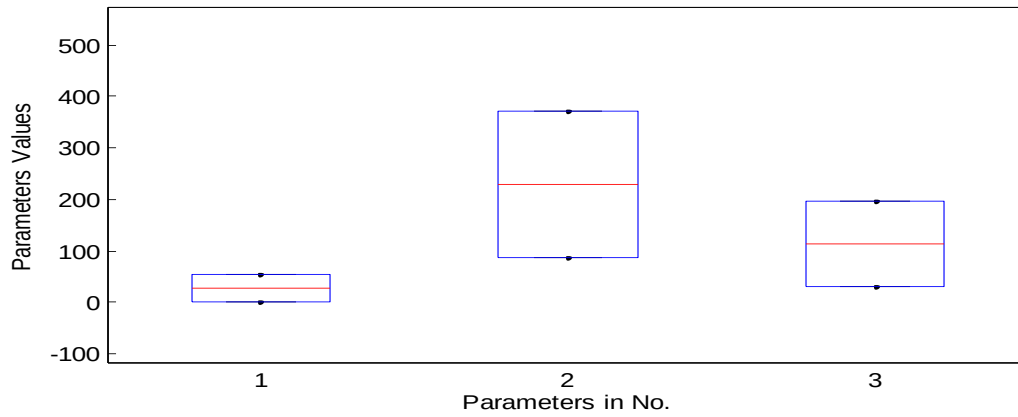
By Spectrum Analysis

Table 5.12(a): Parameters of word 'KITE'

	S1	S2	S3	S4	Avg	Stdev
MEAN	6.15	5.84	5.15	5.89	5.76	0.43
MEDIAN	3.46	3.33	2.75	3.3	3.21	0.31
STD	7.34	7.136	6.86	7.01	7.09	0.20
RMS	9.58	9.22	8.58	9.15	9.13	0.41
MAX	56.48	53.68	54.3	54	54.62	1.27
MIN	0.0021	0.0012	0.012	0.0014	0.004	0.005
SLOP	5.4	5.7	5.4	5.3	5.45	0.17
WIDTH	59	55	50	53	54.25	3.77
P.FREQ(Hz)	320.41	316.41	500.61	343.75	370.30	87.71
P.AMP(dB)	-32.69	-37.75	-27.15	-26.69	-31.07	5.22
T.POW(dB)	-18.77	-19.91	-11.63	-12.06	-15.59	4.36
THD(%)	76.29	92.79	52.26	60.47	70.45	17.92
THD+N (%)	195.26	240.83	183.56	171.12	197.69	30.40
IMD(%)	20.76	103.6	49.73	39.16	53.31	35.60
SNR(dB)	-5.813	-7.63	-5.276	-4.304	-5.76	1.40



Graph 5.12(a): Insignificant Parameters from table 5.12(a)

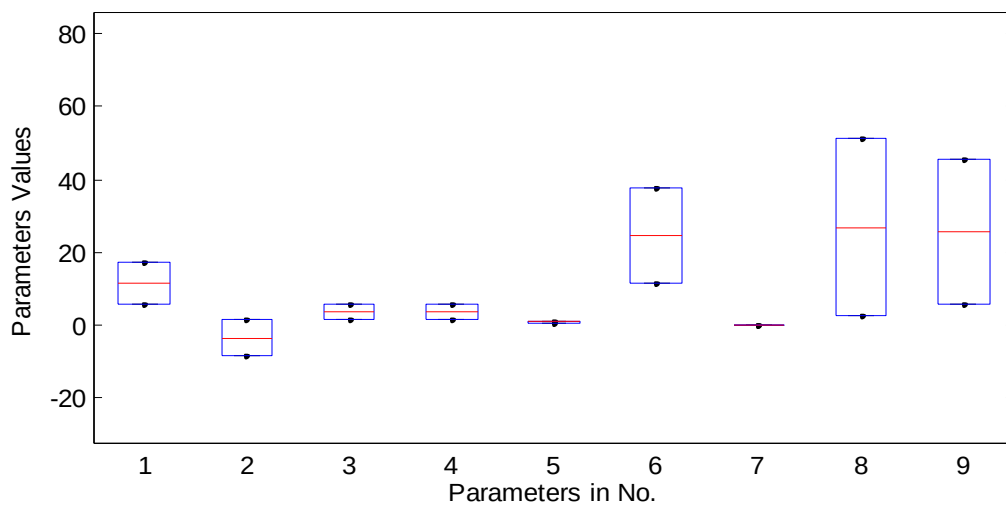


Graph 5.12(b): Significant Parameters from table 5.12(a)

By LPC Spectrum Analysis

Table 5.12(b): Parameters of word 'KITE'

	S1	S2	S3	S4	Avg	Stdev
MAX	19.755	16.2798	9.8783	23.7319	17.41	5.87
MIN	-6.8499	-9.9344	-7.7097	-10.0407	-8.63	1.60
STD	5.5334	6.6231	3.8879	7.0486	5.77	1.41
RMS	5.5118	6.5973	3.8727	7.021	5.75	1.40
SLOP	1.36	1.63	0.7	1.44	1.28	0.40
WIDTH	46	22	37	46	37.75	11.32
FRQ	0.0136	0.0117	0.0136	0.0156	0.014	0.002
MAXMAG	50.57	54.6	48.02	51.47	51.17	2.72
A.TION(dB)	47.1	39.5	43.1	52.3	45.50	5.49



Graph 5.12(c): Significant Parameters from table 5.12(b)

13 NIGHT

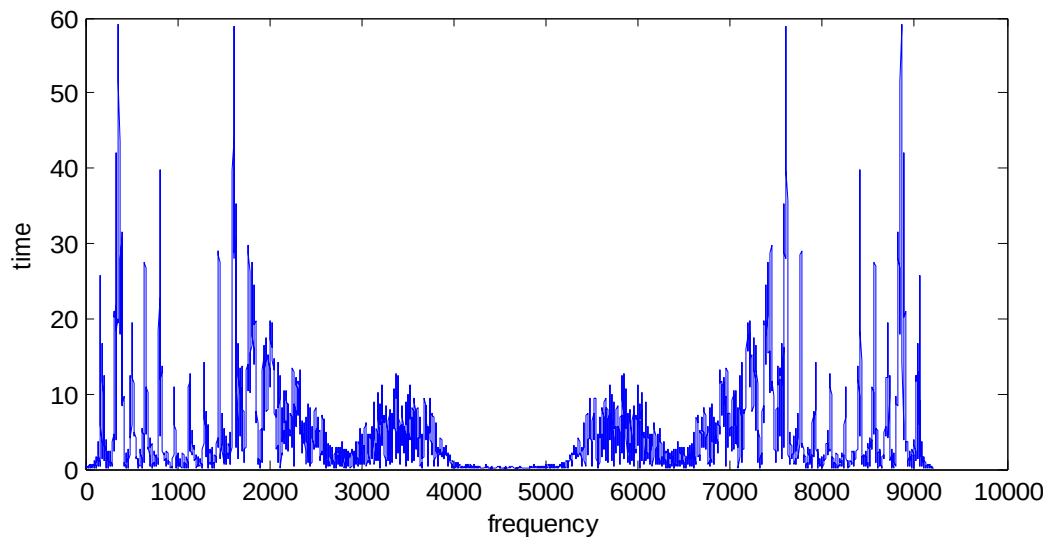
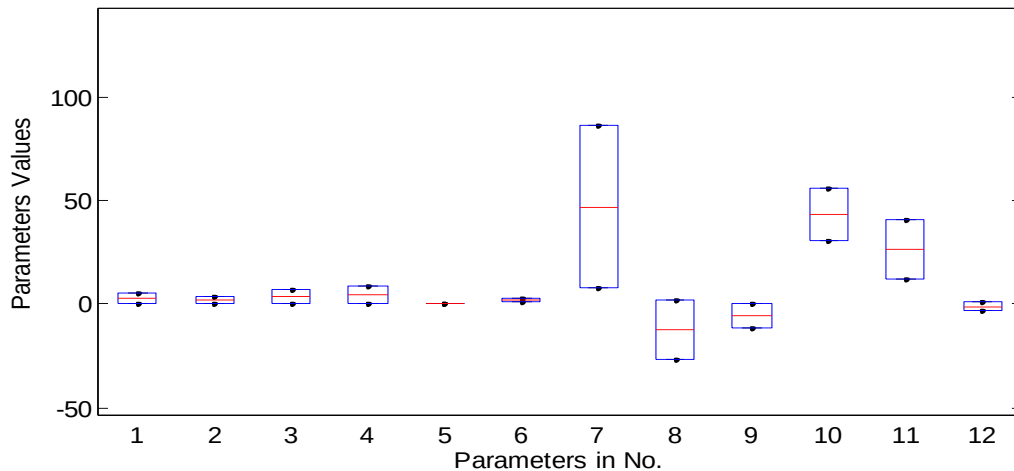


Figure 5.13: Spectrum of 'NIGHT' word

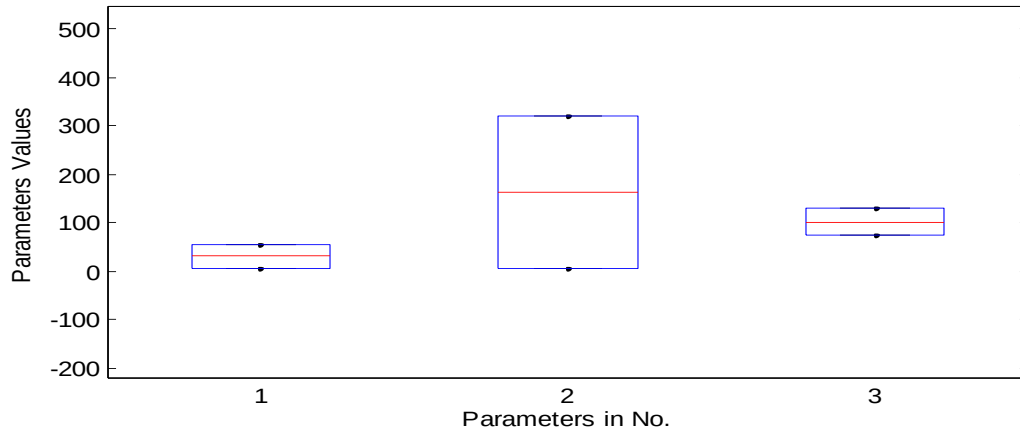
By Spectrum Analysis

Table 5.13(a): Parameters of word 'NIGHT'

	S1	S2	S3	S4	Avg	Stdev
MEAN	6.49	5.64	4.93	5.49	5.64	0.65
MEDIAN	4.18	3.24	2.55	3.64	3.40	0.69
STD	7.03	7.25	6.97	6.99	7.06	0.13
RMS	9.57	9.19	8.54	9.01	9.08	0.43
MAX	46.51	57.68	59.03	58.01	55.31	5.89
MIN	0.101	0.013	0.014	0.012	0.035	0.044
SLOP	1.308	3.58	2.2	3.47	2.64	1.09
WIDTH	98	84	80	85	86.75	7.80
P. FREQ(Hz)	328.12	318.36	320.39	312.56	319.86	6.43
P. AMP(dB)	-29.5	-25.09	-25.93	-26.87	-26.85	1.91
T. POW(dB)	-10.61	-11.61	-11.53	-11.55	-11.33	0.48
THD(%)	91.84	47.07	19.1	66.22	56.06	30.72
THD+N (%)	167.62	173.24	16.99	158.57	129.11	74.99
IMD(%)	46.26	43.07	23.66	50.9	40.97	11.98
SNR(dB)	-2.55	-2.493	-4.506	-4.004	-3.39	1.02



Graph 5.13(a): Insignificant Parameters from table5.13(a)

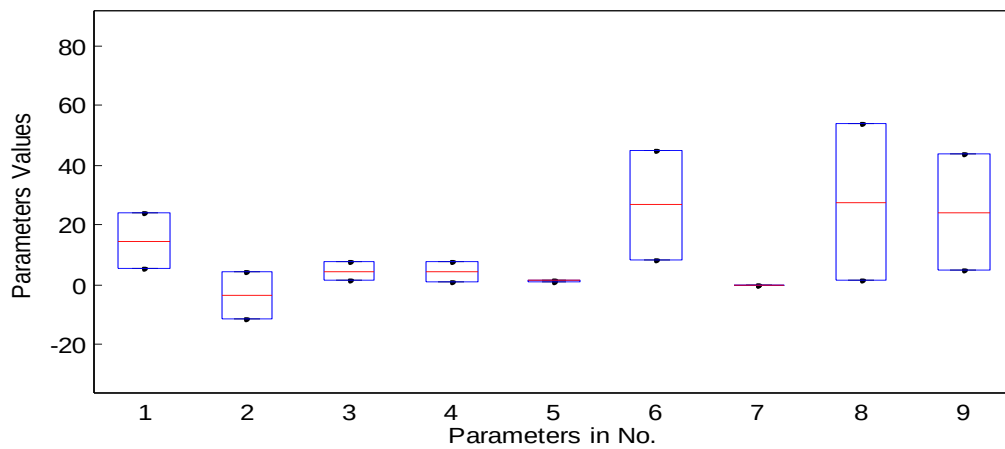


Graph 5.13(b): Significant Parameters from table5.13(a)

By LPC Spectrum Analysis

Table 5.13(b): Parameters of word 'NIGHT'

	S1	S2	S3	S4	Avg	Stdev
MAX	21.4354	24.4759	31.5679	19.3024	24.20	5.35
MIN	-9.1305	-16.871	-6.9392	-12.419	-11.34	4.32
STD	6.9338	9.2503	6.3253	7.5533	7.52	1.26
RMS	6.9067	9.2142	6.3006	7.5238	7.49	1.26
SLOP	1.09	1.13	3.12	0.92	1.57	1.04
WIDTH	48	53	34	46	45.25	8.06
FRQ	0.0195	0.0136	0.025	0.0195	0.019	0.005
MAXMAG	52.91	56.38	52.98	53.26	53.88	1.67
A.TION(dB)	50.1	40.5	40	44.2	43.70	4.66



Graph 5.13(c): Significant Parameters from table 5.13(b)

14 PUNJAB

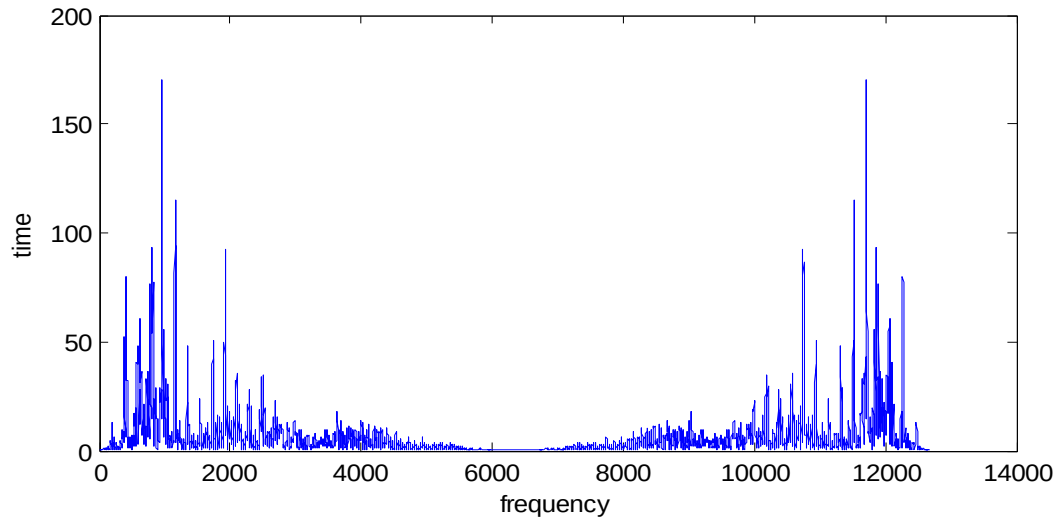
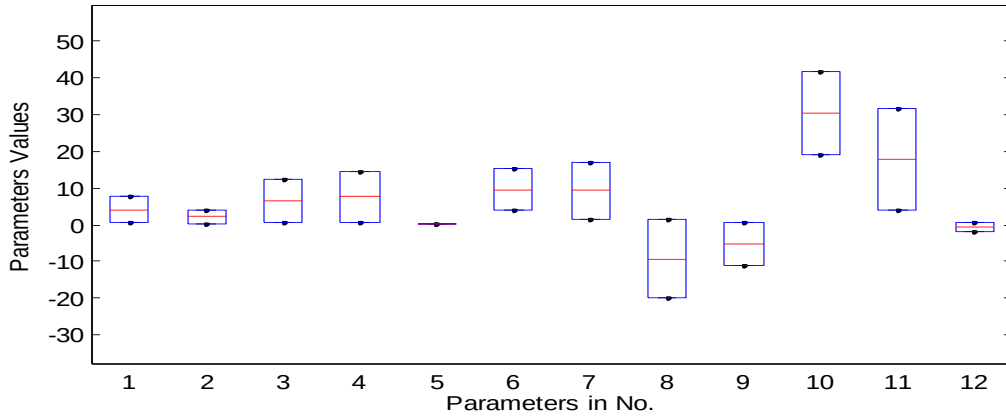


Figure 5.14: Spectrum of 'PUNJAB' word

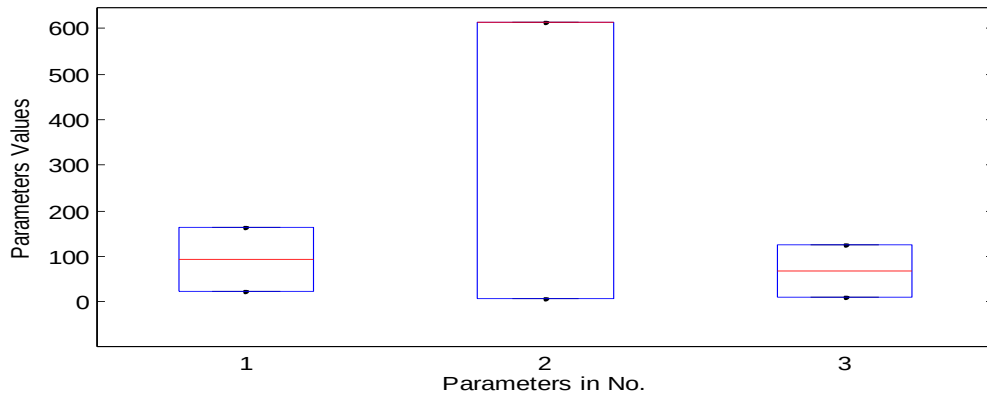
By Spectrum Analysis

Table 5.14(a): Parameters of word 'PUNJAB'

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	8.092	6.9535	8.2334	7.4864	7.7158	7.70	0.51
MEDIAN	4.3234	3.6471	4.3288	3.7017	3.9953	4.00	0.33
STD	12.9039	11.4799	12.6089	11.9465	12.3715	12.26	0.56
RMS	15.2309	13.4212	15.0586	14.0979	14.58	14.48	0.74
MAX	192.779	169.9478	171.3627	130.462	156.37	164.18	22.92
MIN	0.0075	0.0056	0.0022	0.0168	0.0111	0.009	0.006
SLOP	13.77	22.2	13.12	12.87	14.53	15.30	3.91
WIDTH	18	15	17	19	17	17.20	1.48
P.FREQ(Hz)	617.19	607.42	623.05	613.28	609.38	614.06	6.27
P.AMP(dB)	-19.76	-19.62	-21	-22.14	-17.93	-20.09	1.58
T.POW(dB)	-10.05	-11.17	-11.68	-11.73	-10.14	-10.95	0.81
THD(%)	23.08	72.16	27.68	44.4	41.62	41.79	19.22
THD+N (%)	126.83	133.29	114.16	137.86	111.63	124.75	11.55
IMD(%)	24.78	31.56	34.26	35.63	32.26	31.70	4.19
SNR(dB)	-2.065	-2.49	-1.151	-2.789	-0.956	-1.89	0.81



Graph 5.14(a): Insignificant Parameters from table 5.14(a)

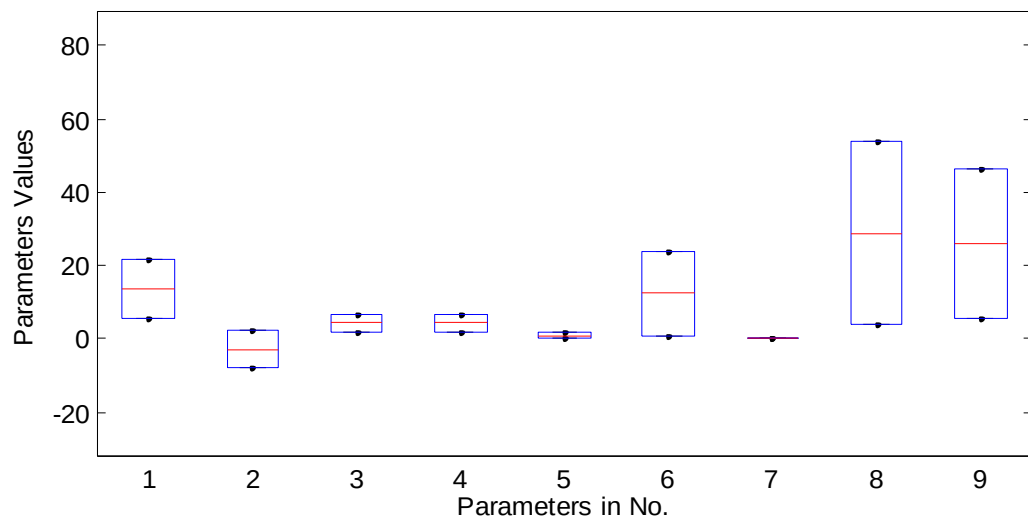


Graph 5.14(b): Significant Parameters from table 5.14(a)

By LPC Spectrum Analysis

Table 5.14(b): Parameters of word 'PUNJAB'

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	26.9989	19.7329	15.4324	18.2016	28.1788	21.71	5.60
MIN	-8.9927	-5.6917	-5.4369	-9.3569	-9.6334	-7.82	2.08
STD	8.3437	6.1413	4.195	7.4168	8.3125	6.88	1.75
RMS	8.3111	6.1173	4.1786	7.3877	8.2801	6.85	1.74
SLOP	2	1.58	1.6	1.57	1.64	1.68	0.18
WIDTH	23	24	24	23	25	23.80	0.84
FRQ	0.0117	0.0117	0.0156	0.0097	0.0097	0.012	0.002
MAXMAG	57.06	52.94	47.16	54.94	56.93	53.81	4.08
A.TION(dB)	45.5	50.1	41.3	53.7	41.5	46.42	5.43



Graph 5.14(c): Significant Parameters from table 5.14(b)

15 INDIA

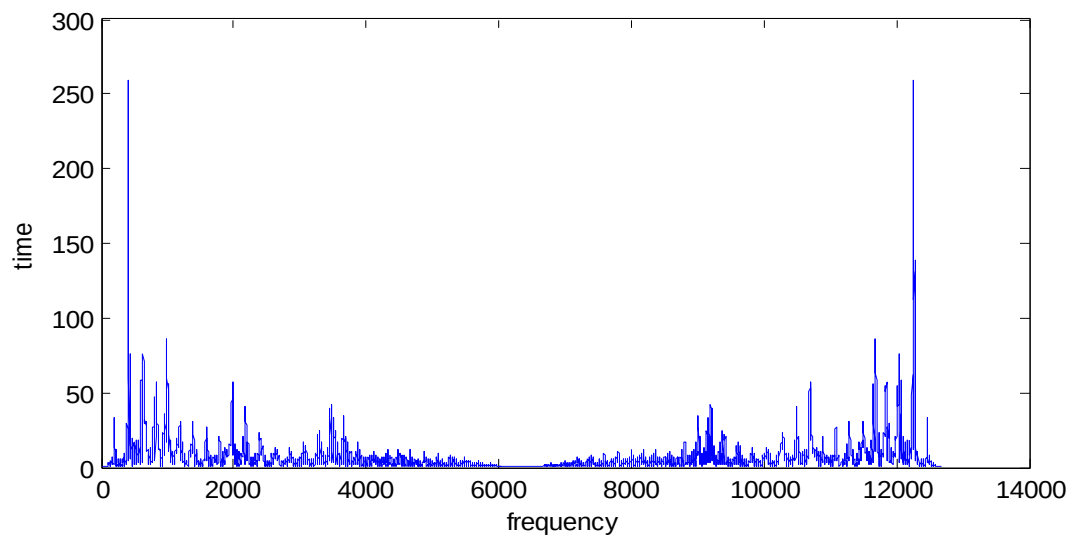
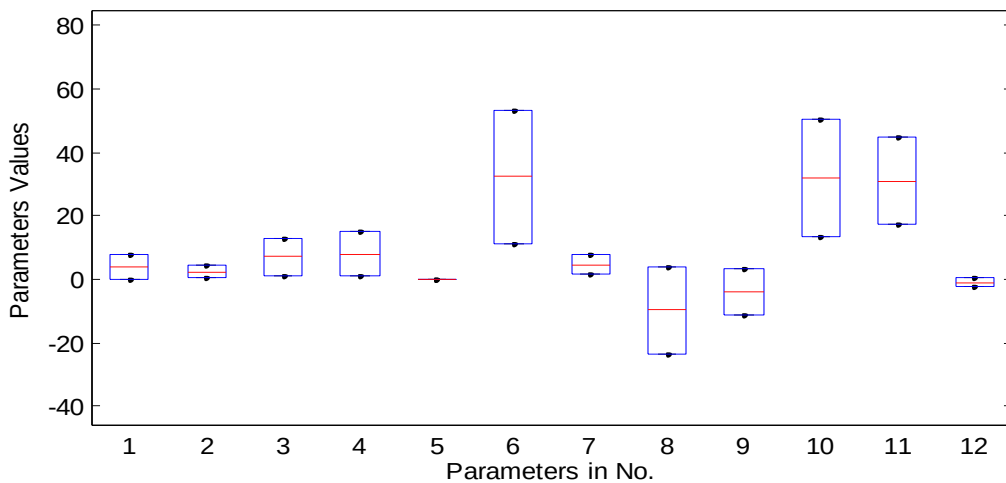


Figure 5.15: Spectrum of 'INDIA' word

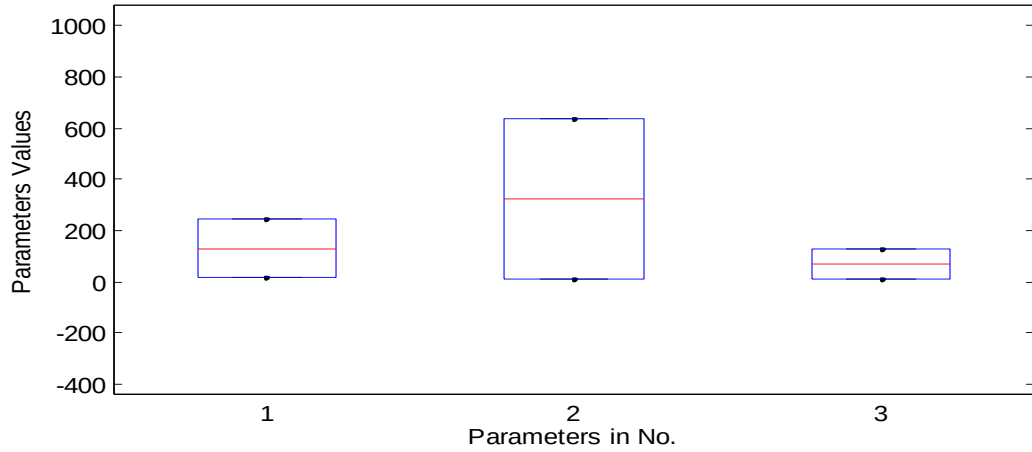
By Spectrum Analysis

Table 5.15(a): Parameters of word ‘INDIA’

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	7.8185	7.698	8.1031	7.4089	7.8323	7.77	0.25
MEDIAN	4.7941	4.1972	4.5158	3.9518	4.3996	4.37	0.32
STD	11.102	12.8942	13.9968	13.836	12.9425	12.95	1.15
RMS	13.5784	15.0169	16.1727	15.6943	15.1275	15.12	0.98
MAX	214.8734	259.3968	250.7558	244.479	253.2753	244.56	17.43
MIN	0.0115	0.0094	0.0149	0.0103	0.0057	0.010	0.003
SLOP	60.18	36.72	55.75	65.48	48.48	53.32	11.18
WIDTH	7	7	7	9	10	8.00	1.41
P.FREQ(Hz)	637.81	623.05	621.09	644.53	648.48	634.99	12.42
P.AMP(dB)	-20.69	-23.71	-29.3	-19.84	-22.99	-23.31	3.71
T.POW(dB)	-10.06	-11.74	-16.65	-8.98	-9.79	-11.44	3.08
THD(%)	60.32	68.16	47.39	38.72	37.86	50.49	13.38
THD+N (%)	121.38	122.17	132.39	125.75	141.03	128.54	8.22
IMD(%)	35.95	62.68	53.35	19.65	53	44.93	17.11
SNR(dB)	-2.238	-1.739	-2.43	-2.29	-2.98	-2.34	0.44



Graph 5.15(a): Insignificant Parameters from table 5.15(a)

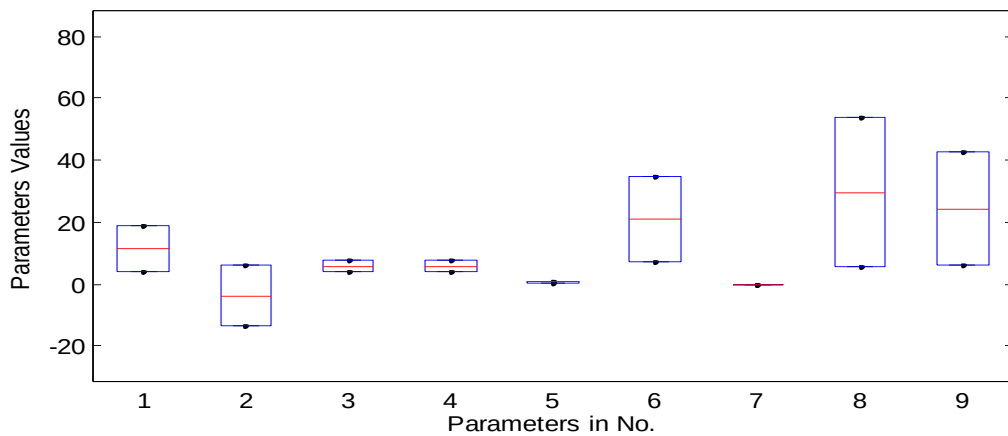


Graph 5.15(b): Significant Parameters from table 5.15(a)

By LPC Spectrum Analysis

Table 5.15(b): Parameters of word ‘INDIA’

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	19.4535	12.9493	23.755	21.1006	15.9696	18.65	4.25
MIN	-11.7369	-11.2155	-24.5812	-9.8064	-10.7733	-13.62	6.17
STD	8.5981	4.5716	14.006	6.6807	4.7413	7.72	3.88
RMS	8.5645	4.5538	13.9513	6.6545	4.7229	7.69	3.86
SLOP	0.67	0.75	0.93	1.35	1.18	0.98	0.29
WIDTH	35	43	35	38	23	34.80	7.36
FRQ	0.0136	0.0273	0.0175	0.0156	0.0312	0.021	0.008
MAXMAG	56.5	48.39	61.73	52.64	49.18	53.69	5.52
A.TION(dB)	50.1	35.1	44.9	45.1	37.2	42.48	6.19



Graph5.15(c) Significant Parameters from table5.15(b)

16 PAKSITAN

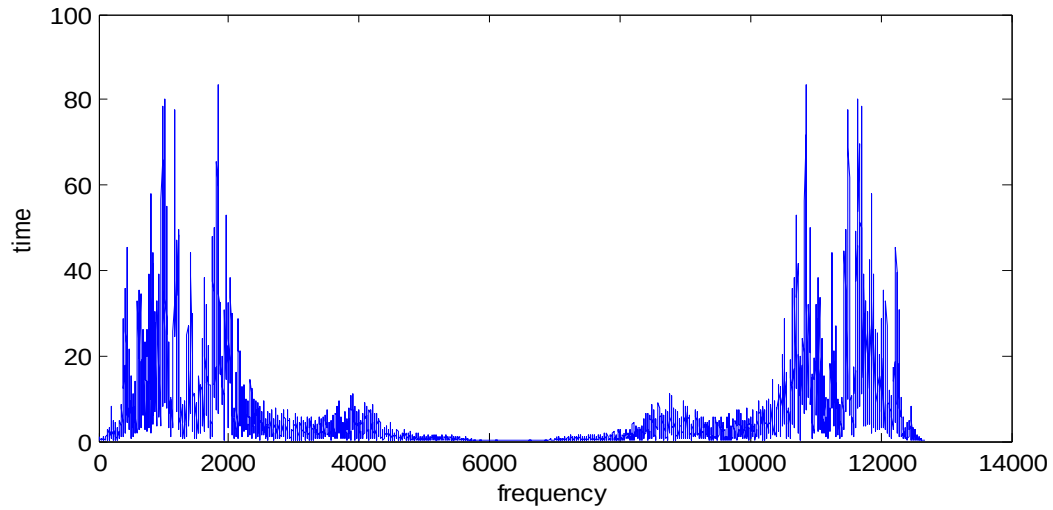
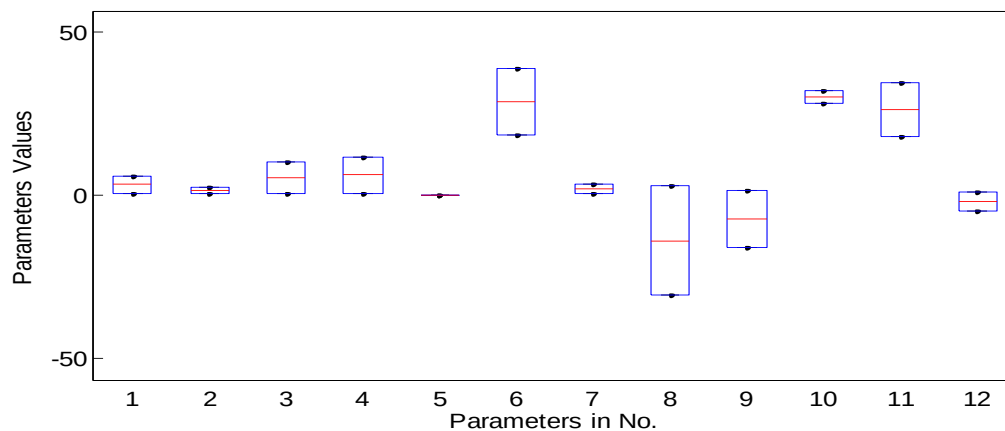


Figure 5.16: Spectrum of 'PAKSITAN' word

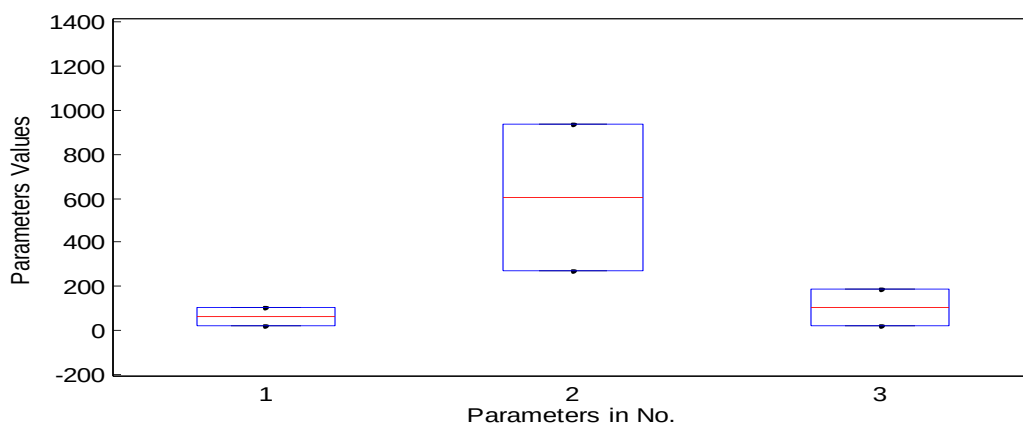
By Spectrum Analysis

Table 5.16(a): Parameters of word 'PAKSITAN'

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	5.3751	5.6482	6.5139	6.4841	6.1034	6.02	0.51
MEDIAN	2.2245	2.1453	2.5895	2.7636	2.7033	2.49	0.28
STD	8.9878	10.1896	10.4218	9.848	10.3319	9.96	0.58
RMS	10.4721	11.65	12.2897	11.7906	11.9996	11.64	0.70
MAX	88.5671	126.4908	83.6161	80.9285	133.5073	102.62	25.26
MIN	0.0018	0.0101	0.0081	0.0108	0.0047	0.007	0.004
SLOP	15.47	42.93	35.65	33.65	66.82	38.90	18.59
WIDTH	4	3	3	3	3	3.20	0.45
P.FREQ(Hz)	1107.42	1140.62	1154.3	628.91	642.58	934.77	273.54
P.AMP(dB)	-27.42	-30.33	-28.4	-33.25	-34.26	-30.73	2.97
T.POW(dB)	-13.59	-16.1	-15.59	-17.11	-17.41	-15.96	1.52
THD(%)	5.179	5.089	5.73	56.68	68.85	28.31	31.75
THD+N (%)	178.07	178.79	153.55	197.12	218.25	185.16	24.12
IMD(%)	26.48	30.98	10.26	48.2	55.37	34.26	17.94
SNR(dB)	-5.012	-5.043	-3.261	-5.191	-6.339	-4.97	1.10



Graph 5.16(a): Insignificant Parameters from table 5.16(a)

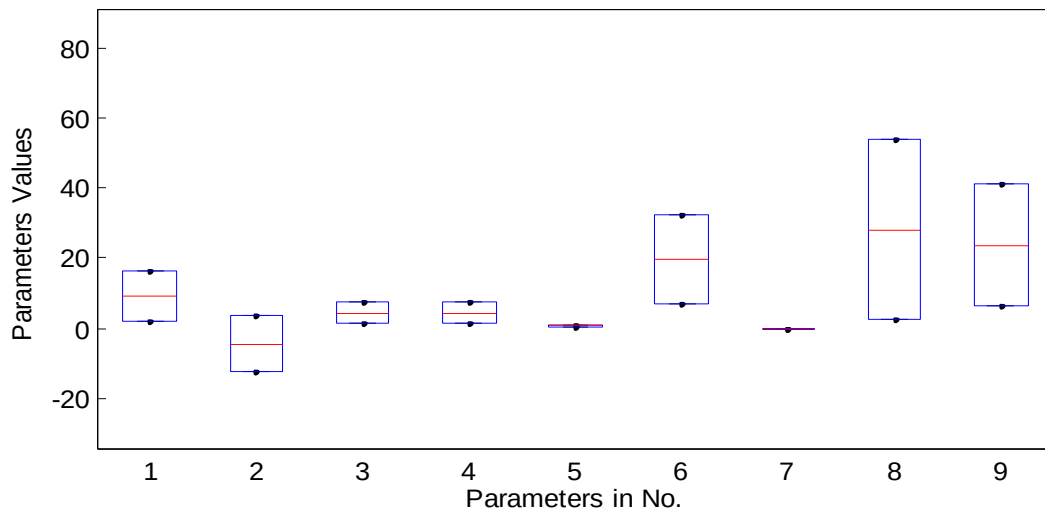


Graph 5.16(b): Significant Parameters from table 5.16(a)

By LPC Spectrum Analysis

Table 5.16(b): Parameters of word ‘PAKSITAN’

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	18.9681	16.1607	14.3001	18.1053	15.4502	16.60	1.92
MIN	-9.899	-14.7	-14.821	-15.128	-7.8187	-12.47	3.38
STD	8.9155	7.733	6.8723	8.3618	5.2878	7.43	1.42
RMS	8.8806	7.703	6.8456	8.3293	5.2671	7.41	1.41
SLOP	0.97	1.16	1.16	1.23	1.13	1.13	0.10
WIDTH	21	35	34	39	34	32.60	6.80
FRQ	0.0136	0.0117	0.042	0.0117	0.0136	0.019	0.013
MAXMAG	57.59	54.29	51.04	54.43	51.46	53.76	2.65
A.TION(dB)	47.8	36.8	34.6	37.7	48.3	41.04	6.50



Graph 5.16(c): Significant Parameters from table 5.16(b)

17 MINISTER

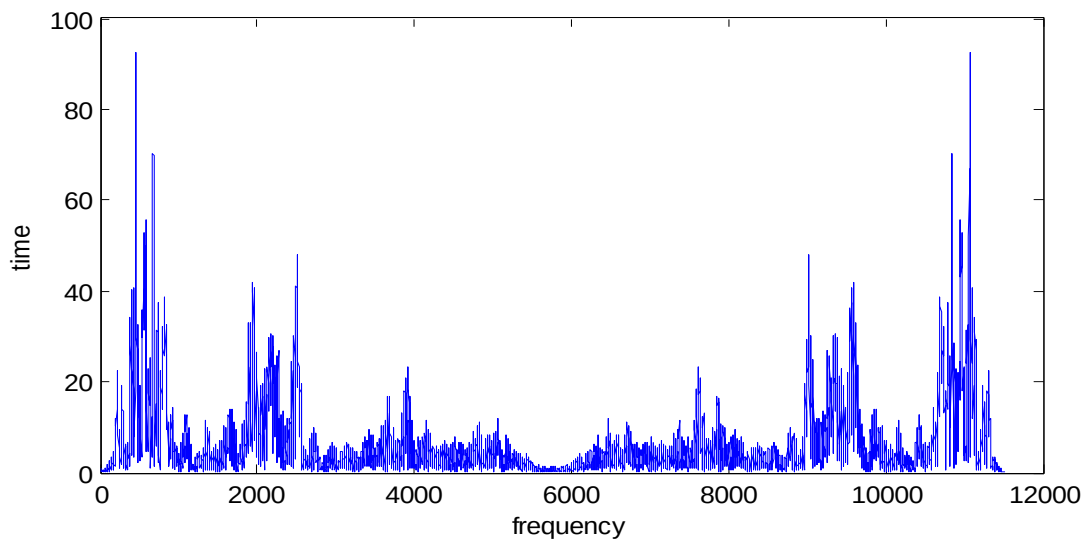
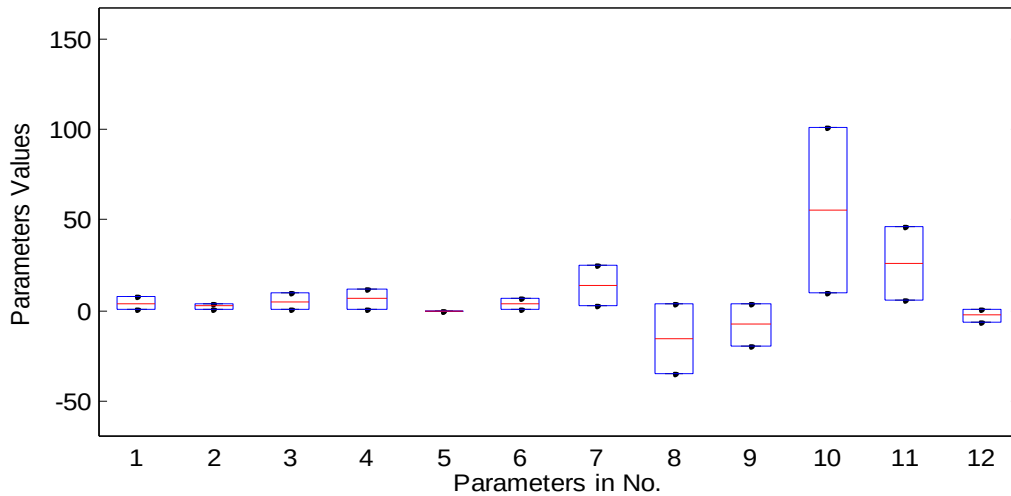


Figure 5.17: Spectrum of 'MINISTER' word

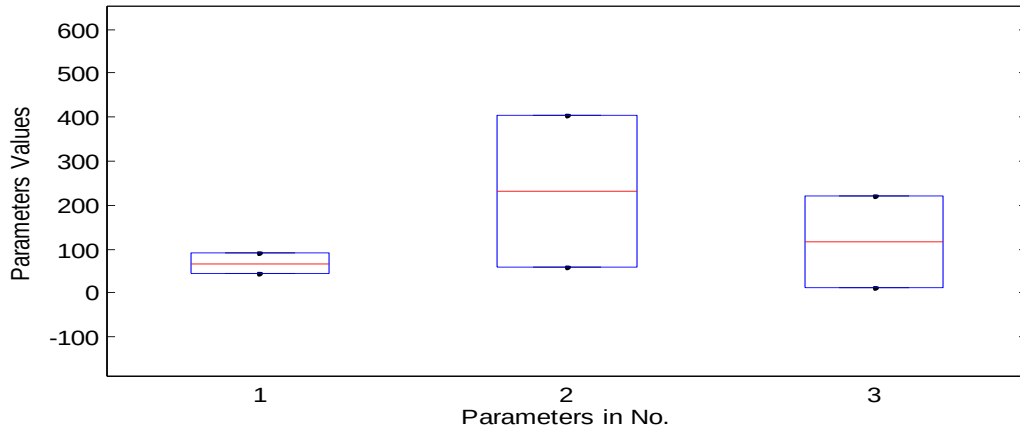
By Spectrum Analysis

Table 5.17(a): Parameters of word ‘MINISTER’

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	6.9174	7.7793	7.1262	7.2908	6.8964	7.20	0.36
MEDIAN	3.8586	4.7954	3.8649	3.2961	3.2961	3.82	0.61
STD	8.9423	9.7023	10.4587	9.7439	8.2048	9.41	0.86
RMS	11.3052	12.4356	12.6554	12.3401	11.419	12.03	0.62
MAX	92.3457	109.661	124.0476	113.649	15.294	91.00	43.84
MIN	0.0161	0.0183	0.0062	0.0153	0.0171	0.01	0.00
SLOP	6.35	5.47	7.67	6.36	6.82	6.53	0.80
WIDTH	20	28	24	26	25	24.60	2.97
P.FREQ(Hz)	507.81	359.38	386.72	370.38	392.47	403.35	59.85
P.AMP(dB)	-33.18	-39.99	-30.42	-34.28	-36.27	-34.83	3.58
T. POW(dB)	-20.25	-23.37	-14.25	-18.25	-20.46	-19.32	3.37
THD(%)	90.48	91.96	111.08	103.29	108.59	101.08	9.44
THD+N (%)	206.97	217.49	235.6	215.59	227.48	220.63	11.10
IMD(%)	43.3	48.31	38.58	52.47	49.37	46.41	5.48
SNR(dB)	-5.34	-6.74	-7.44	-6.49	-5.39	-6.28	0.91



Graph 5.17(a): Insignificant Parameters from table 5.17(a)

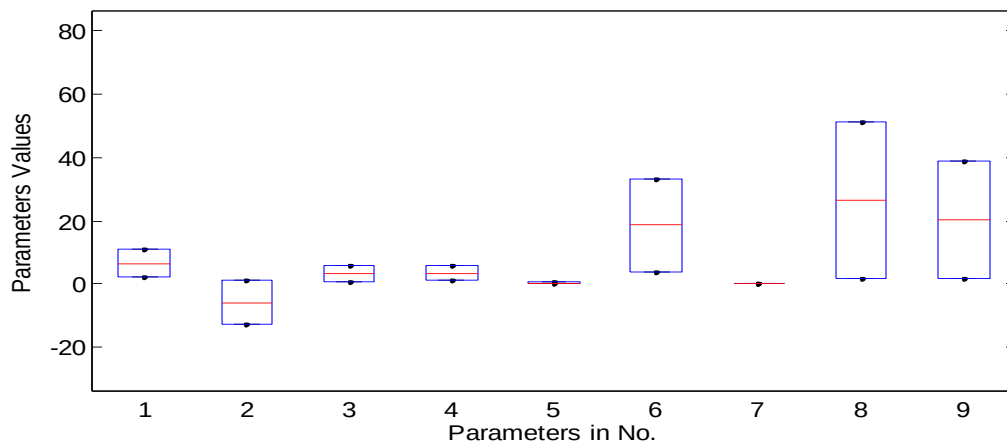


Graph 5.17(b): Significant Parameters from table 5.17(a)

By LPC Spectrum Analysis

Table 5.17(b): Parameters of word ‘MINISTER’

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	9.73	14.095	8.825	9.843	11.732	10.85	2.10
MIN	-14.503	-12.78	-12.177	-11.83	-12.473	-12.75	1.04
STD	6.278	6.432	4.834	4.672	6.384	5.72	0.89
RMS	6.253	6.407	4.815	4.392	6.759	5.73	1.05
SLOP	0.34	0.69	0.88	0.79	0.7	0.68	0.21
WIDTH	39	33	28	32	35	33.40	4.04
FRQ	0.0195	0.0195	0.0175	0.0175	0.0172	0.02	0.00
MAXMAG	52.93	52.92	50.67	50.98	48.29	51.16	1.92
A.TION(dB)	37.8	41.5	37.5	38.9	40.2	39.18	1.68



Graph 5.17(c): Significant Parameters from table 5.17(b)

18 MURDER

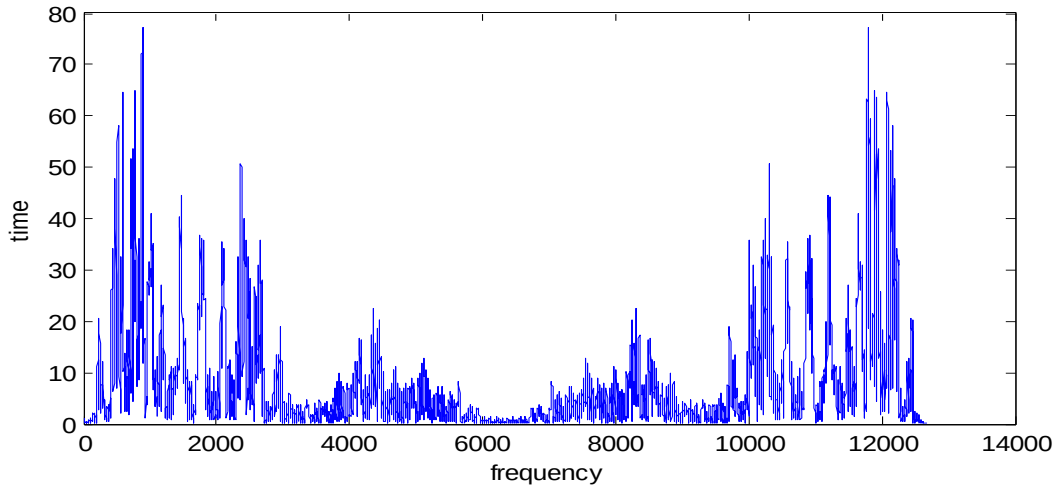
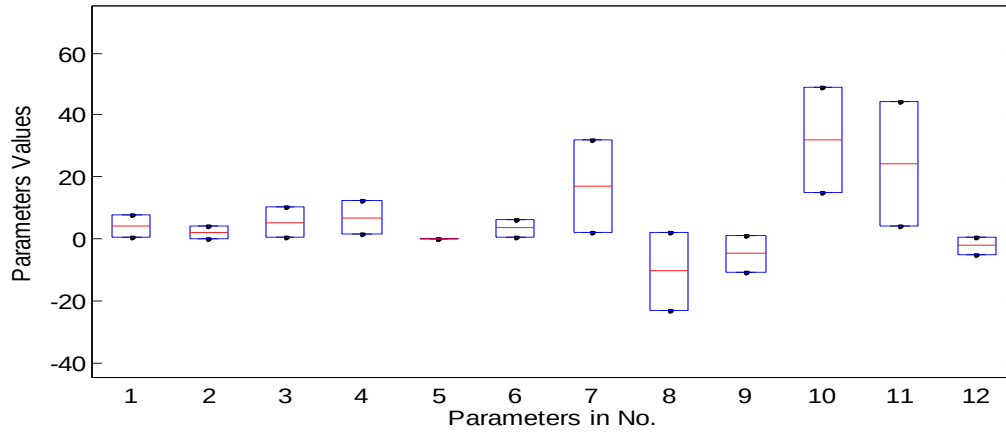


Figure 5.18: Spectrum of 'MURDER' word

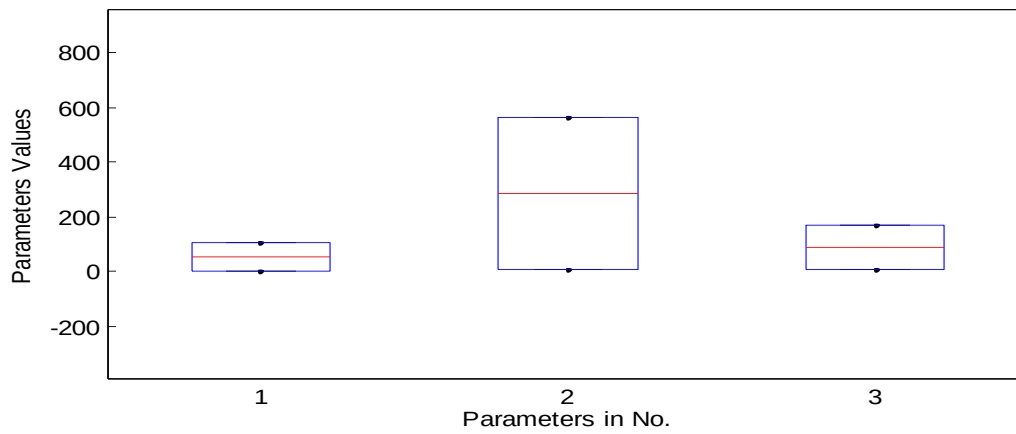
By Spectrum Analysis

Table 5.18(a): Parameters of word 'MURDER'

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	7.897	8.1292	7.2934	8.2934	6.9283	7.71	0.58
MEDIAN	4.0625	4.0742	4.2934	4.0892	3.9547	4.09	0.12
STD	11.1746	10.3077	10.2893	10.2934	9.9384	10.40	0.46
RMS	13.683	13.1272	12.832	11.2034	10.2384	12.22	1.44
MAX	111.6594	97.3704	103.289	107.2934	105.8363	105.09	5.28
MIN	0.0201	0.0077	0.0149	0.0162	0.0173	0.02	0.00
SLOP	7.37	6.29	5.39	5.47	6.2	6.14	0.80
WIDTH	30	32	35	29	33	31.80	2.39
P.FREQ(Hz)	548.83	556.64	574.22	570.23	562.38	562.46	10.23
P.AMP(dB)	-19.65	-23.52	-26.11	-24.49	-22.43	-23.24	2.42
T.POW(dB)	-9.82	-9.48	-12.7	-10.02	-11.27	-10.66	1.33
THD(%)	60.11	57.86	23.12	55.49	49.39	49.19	15.11
THD+N (%)	158.34	186.91	167.95	170.32	164.38	169.58	10.69
IMD(%)	42.61	47.4	39.16	42.65	49.38	44.24	4.10
SNR(dB)	-4.69	-5.43	-4.48	-4.2	-5.71	-4.90	0.64



Graph 5.18(a): Insignificant Parameters from table 5.18(a)

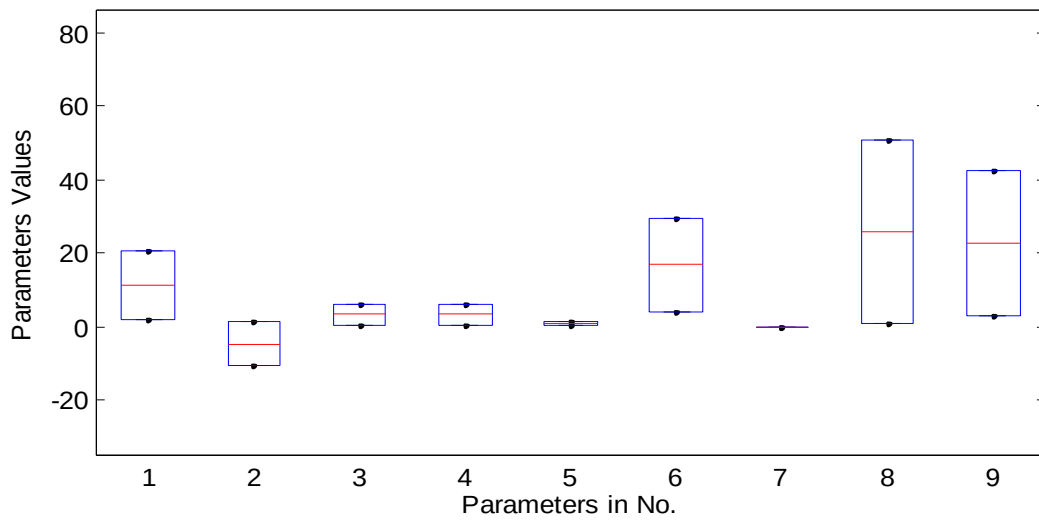


Graph 5.18(b): Significant Parameters from table5.18(a)

By LPC Spectrum Analysis

Table 5.18(b): Parameters of word ‘MURDER’

	S1	S2	S3	S4	S5	Avg	Std
MAX	20.007	24.156	18.257	20.384	21.384	20.84	2.17
MIN	-9.72	-12.086	-11.273	-8.932	-9.382	-10.28	1.34
STD	6.395	6.634	5.283	5.493	6.732	6.11	0.67
RMS	6.37	6.608	5.382	5.639	6.835	6.17	0.63
SLOP	0.89	2.01	1.07	1.78	1.89	1.53	0.51
WIDTH	36	29	24	28	30	29.40	4.34
FRQ	0.0152	0.0156	0.024	0.014	0.0147	0.02	0.00
MAXMAG	52.25	51.51	50.12	50.24	49.39	50.70	1.15
A.TION(dB)	46.3	42.2	38.49	43.83	40.78	42.32	2.97



Graph 5.18(c): Significant Parameters from table 5.18(b)

19 ENGLAND

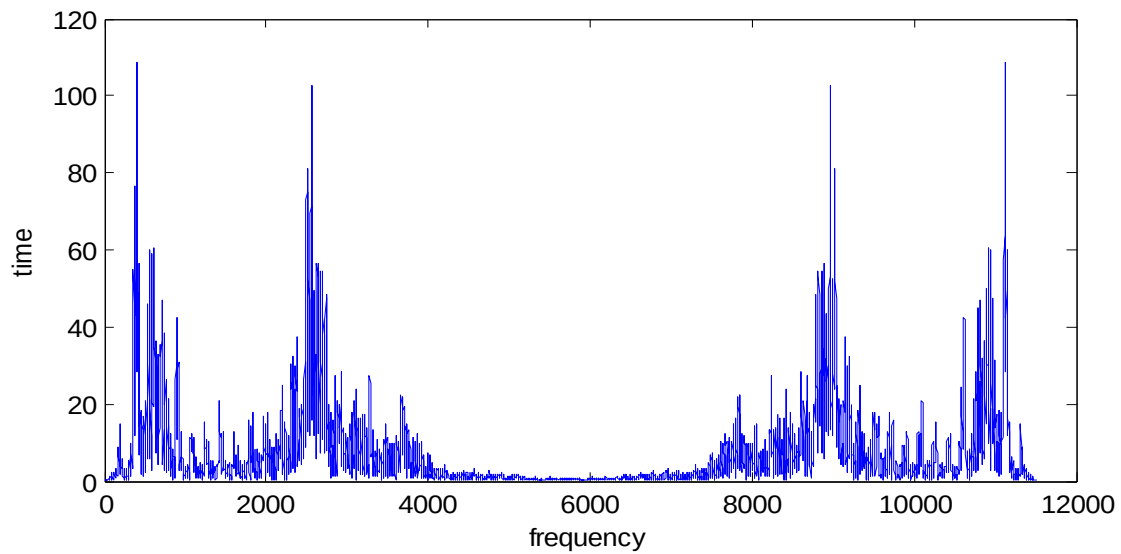
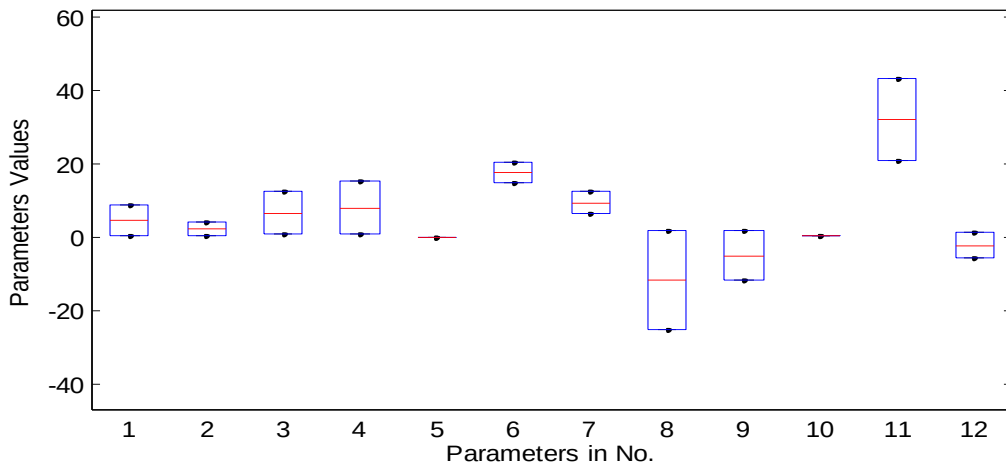


Figure 5.19: Spectrum of 'ENGLAND' word

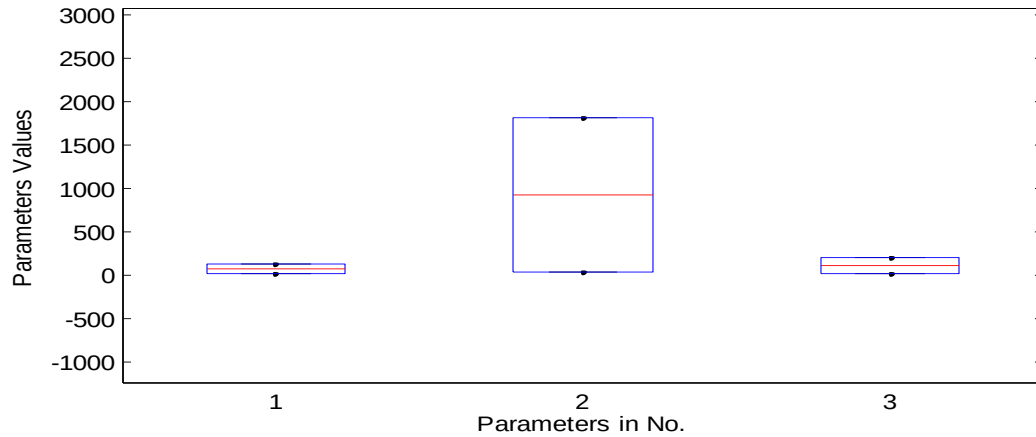
By Spectrum Analysis

Table 5.19(a): Parameters of word ‘ENGLAND’

	S1	S2	S3	S4	S5	S6	Avg	Stdev
MEAN	7.8616	8.9037	8.2067	8.9192	8.871	8.1642	8.49	0.47
MEDIAN	3.501	4.4785	4.09	4.4396	4.1631	3.5465	4.04	0.43
STD	11.7765	13.1259	11.6456	12.7174	12.7582	12.6358	12.44	0.59
RMS	14.159	15.8604	14.2464	15.5329	15.5388	15.0434	15.06	0.72
MAX	108.8505	135.7772	137.7463	133.5188	134.576	111.2596	126.95	13.19
MIN	0.0261	0.0046	0.0074	0.0085	0.0144	0.0055	0.011	0.008
SLOP	8.044	25.86	9.5	30.31	42.75	6.35	20.47	14.81
WIDTH	19	9	14	7	4	20	12.17	6.55
P.FREQ(Hz)	1773.44	1867.19	1791.02	1827.27	1765.62	1854.38	1813.15	42.77
P.AMP(dB)	-28.72	-25.2	-24.19	-26.36	-24.01	-23.7	-25.36	1.91
T.POW(dB)	-13.04	-12.83	-9.36	-11.84	-13.5	-11.07	-11.94	1.54
THD(%)	1.009	0.408	0.52	0.43	0.23	0.251	0.47	0.28
THD+N (%)	226.82	174.98	177.22	209.04	218.89	164.86	195.30	26.10
IMD(%)	13.67	50.22	66.28	50.71	21.63	54.92	42.91	20.55
SNR(dB)	-7.104	-4.861	-4.97	-6.89	-6.805	-4.342	-5.83	1.23



Graph 5.19(a): Insignificant Parameters from table 5.19(a)

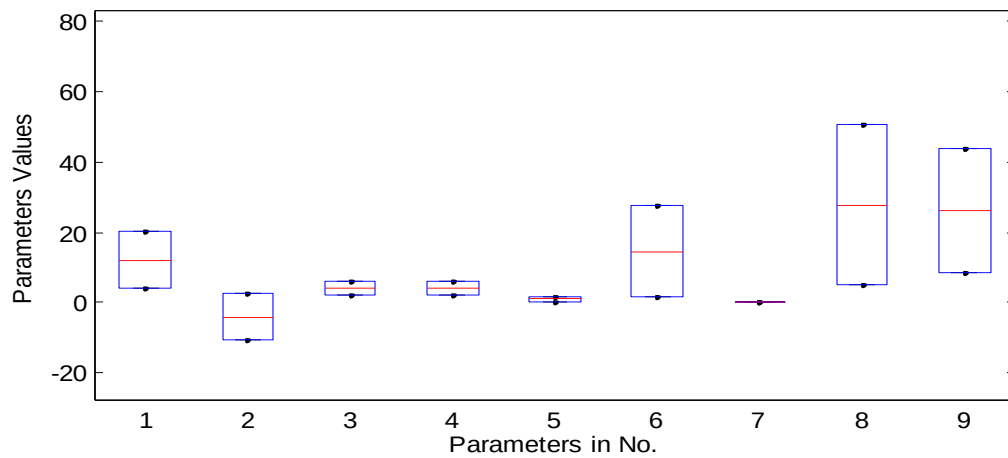


Graph 5.19(b): Significant Parameters from table 5.19(a)

By LPC Spectrum Analysis

Table 5.19(b): Parameters of word 'ENGLAND'

	S1	S2	S3	S4	S5	S6	Avg	Stdev
MAX	18.7474	22.0184	17.6131	20.629	14.9854	27.0975	20.18	4.17
MIN	-8.9378	-7.2488	-12.945	-13.078	-8.9537	-12.925	-10.68	2.60
STD	3.7917	4.5813	7.7505	8.7342	3.8693	8.0737	6.13	2.29
RMS	3.777	4.5633	7.7202	8.7	3.8542	8.0421	6.11	2.28
SLOP	1.84	2.08	1.46	1.76	1.59	1.89	1.77	0.22
WIDTH	26	28	28	28	26	30	27.67	1.51
FRQ	0.097	0.046	0.0156	0.0156	0.095	0.0117	0.047	0.040
MAXMAG	45.19	46.99	54.76	56.39	45.39	54.8	50.59	5.25
A.TION(dB)	34.3	53.7	44	44.1	33.7	52.7	43.75	8.60



Graph 5.19(c): Significant Parameters from table 5.19(b)

20 CANADA

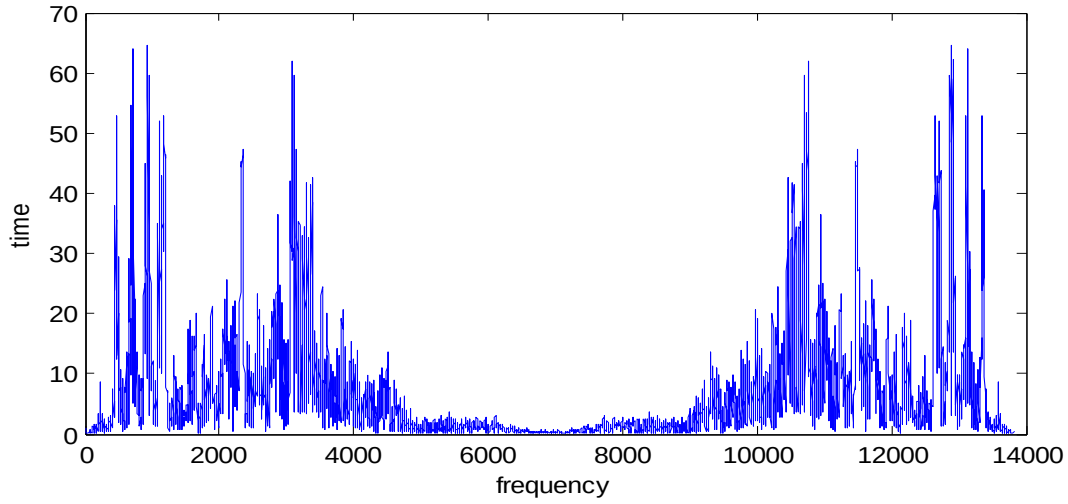
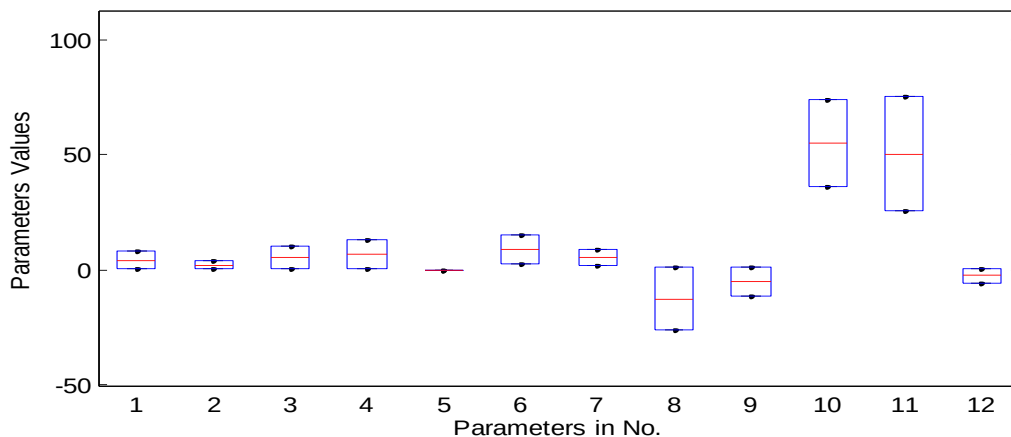


Figure 5.20: Spectrum of 'CANADA' word

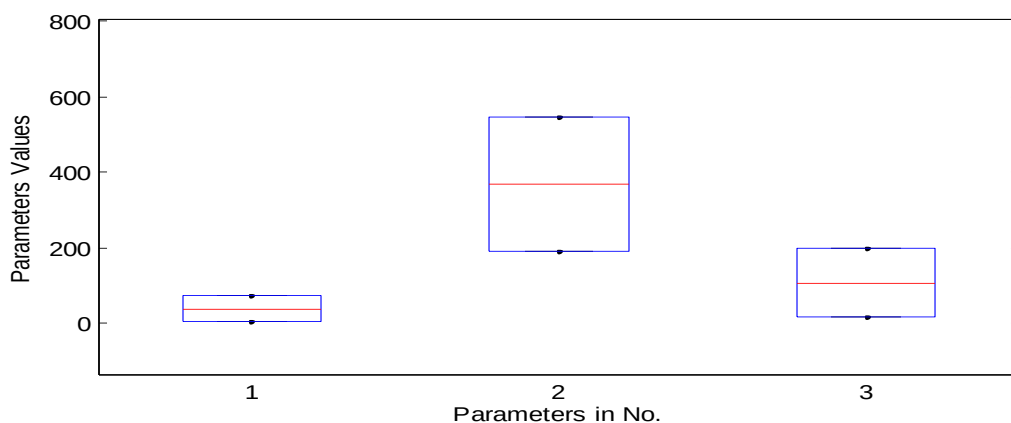
By Spectrum Analysis

Table 5.20(a): Parameters of word 'CANADA'

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	8.0483	7.9242	7.4188	8.3062	8.1131	7.96	0.33
MEDIAN	4.2729	3.8519	3.9476	4.6282	4.1333	4.17	0.31
STD	10.1589	10.1633	9.5809	10.385	10.9533	10.25	0.49
RMS	12.9603	12.8871	12.1171	13.2979	13.6304	12.98	0.57
MAX	76.383	71.7696	64.5024	72.578	72.1317	71.47	4.31
MIN	0.0093	0.0098	0.0044	0.0058	0.0029	0.006	0.003
SLOP	16.25	10.68	15.97	16.16	18.9	15.59	3.00
WIDTH	12	10	9	7	7	9.00	2.12
P.FREQ(Hz)	673.83	421.88	273.44	681.64	679.69	546.10	188.62
P. AMP(dB)	-26.86	-26.21	-27.76	-25.17	-24.58	-26.12	1.28
T. POW(dB)	-10.78	-12.37	-12.95	-10.47	-9.69	-11.25	1.36
THD(%)	48.51	46.76	130.02	53.55	91.6	74.09	36.26
THD+N (%)	180.31	189.23	184.55	211.42	217.85	196.67	16.85
IMD(%)	74.91	40.95	109.44	87.97	63.48	75.35	25.72
SNR(dB)	-5.121	-5.54	-5.32	-6.503	-6.76	-5.85	0.74



Graph 5.20(a): Insignificant Parameters from table 5.20(a)

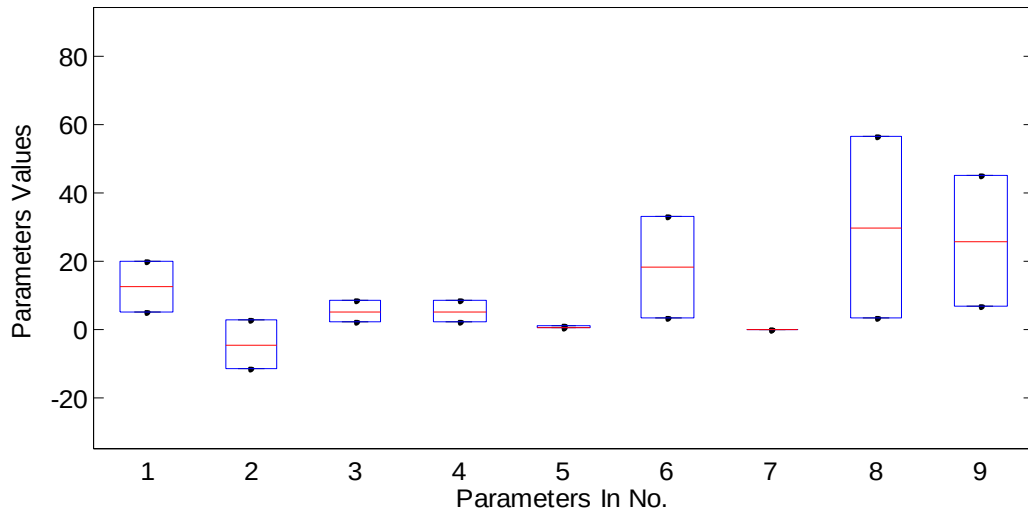


Graph 5.20(b): Significant Parameters from table 5.20(a)

By LPC Spectrum Analysis

Table 5.20(b): Parameters of word 'CANADA'

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	10.8134	20.9283	21.137	25.0097	20.2957	19.64	5.27
MIN	-11.0999	-14.8313	-14.0033	-8.4709	-10.1001	-11.70	2.67
STD	5.1529	10.5021	9.482	8.3508	8.517	8.40	2.01
RMS	5.1329	10.4611	9.4449	8.3181	8.4837	8.37	2.00
SLOP	0.62	0.73	0.75	1.54	0.93	0.91	0.37
WIDTH	27	35	34	35	35	33.20	3.49
FRQ	0.033	0.0117	0.0117	0.0136	0.0117	0.016	0.009
MAXMAG	50.6	59.54	58.27	56.19	56.62	56.24	3.43
A.TION(dB)	35.3	43.8	43.5	48.6	53.1	44.86	6.64



Graph 5.20(c): Significant Parameters from table 5.20(b)

21 AMERICA

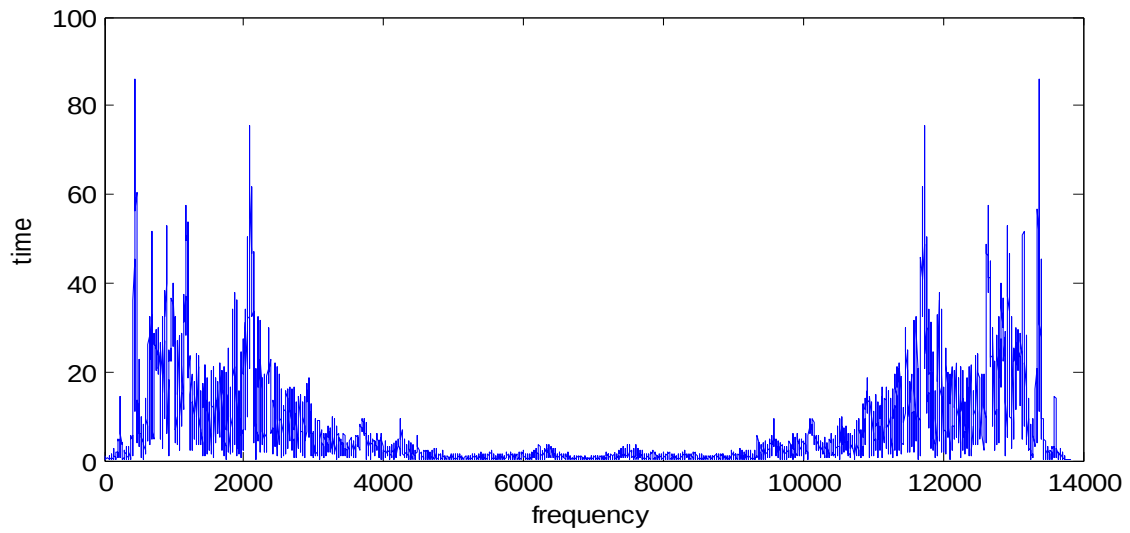
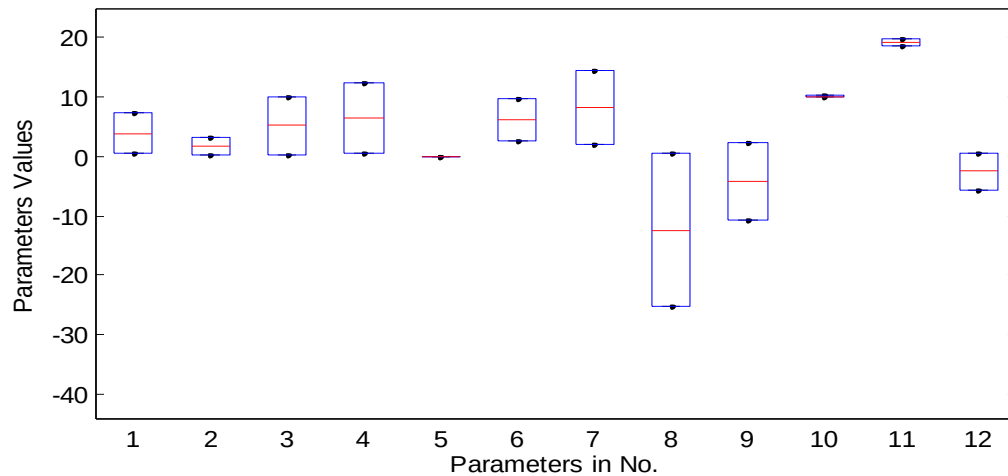


Figure 5.21: Spectrum of 'AMERICA' word

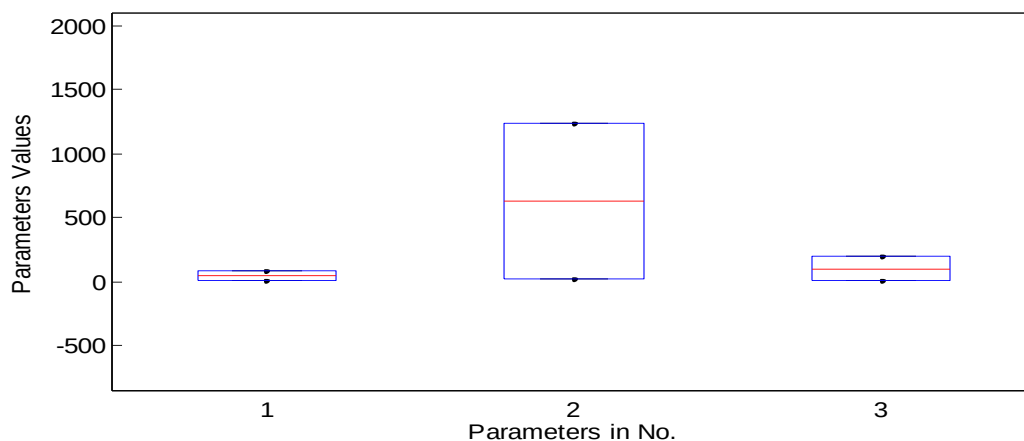
By Spectrum Analysis

Table 5.21(a): Parameters of word ‘AMERICA’

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	7.6264	7.0336	7.4234	7.3741	6.7167	7.23	0.36
MEDIAN	3.4988	2.7434	2.9206	2.8716	2.9109	2.99	0.29
STD	9.8809	9.9632	10.4369	10.2047	9.6361	10.02	0.31
RMS	12.4815	12.1955	12.8073	12.5899	11.7457	12.36	0.41
MAX	76.4183	85.8661	85.576	104.5102	99.3636	90.35	11.39
MIN	0.0048	0.0134	0.0093	0.015	0.0045	0.009	0.005
SLOP	7.21	8.75	9.99	13.9	8.41	9.65	2.57
WIDTH	15	14	11	15	16	14.20	1.92
P.FREQ(Hz)	1216.8	1222.66	1205.08	1265.62	1250	1232.03	24.99
P. AMP(dB)	-25.12	-25.49	-25.13	-25.62	-24.57	-25.19	0.41
T.POW(dB)	-9.8	-9.84	-9.61	-9.63	-14.58	-10.69	2.18
THD(%)	24.74	15.91	4.02	2.158	4.27	10.22	9.77
THD+N (%)	190.79	194.21	204.83	173.37	202.96	193.23	12.56
IMD(%)	4.83	40.96	7.27	6.63	38.93	19.72	18.49
SNR(dB)	-5.566	-5.76	-6.193	-4.78	-6.148	-5.69	0.57



Graph 5.21(a): Insignificant Parameters from table 5.21(a)

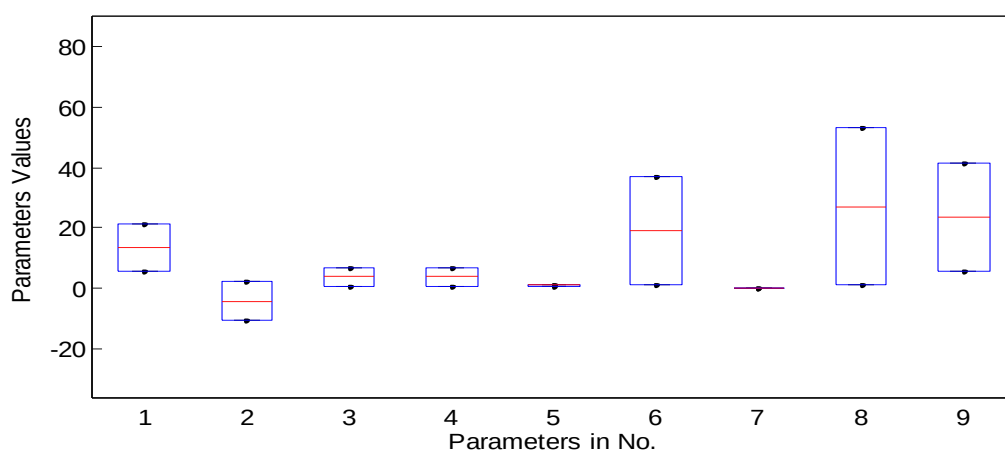


Graph 5.21(b): Significant Parameters from table 5.21(a)

By LPC Spectrum Analysis

Table 5.21(b): Parameters of word ‘AMERICA’

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	19.5781	30.2647	21.8576	20.9158	15.092	21.54	5.52
MIN	-8.6413	-10.797	-8.7087	-10.439	-14.602	-10.64	2.42
STD	6.2311	8.2756	6.2627	7.1963	6.248	6.84	0.90
RMS	6.2068	8.2432	6.2383	7.1682	6.2239	6.82	0.90
SLOP	1.03	2.25	1.43	1.07	1.89	1.53	0.53
WIDTH	39	36	38	36	36	37.00	1.41
FRQ	0.0156	0.0175	0.0195	0.0156	0.021	0.018	0.002
MAXMAG	51.31	54.91	52.91	53.76	52.04	52.99	1.42
A.TION(dB)	46	46.9	37.6	42.4	33	41.18	5.86



Graph 5.21(c): Significant Parameters from table 5.21(b)

22 AUSTRALIA

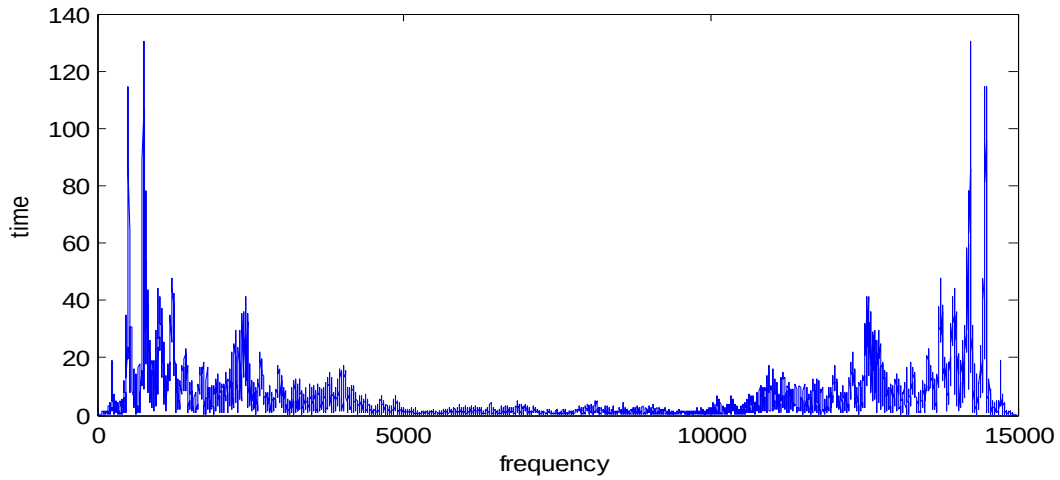
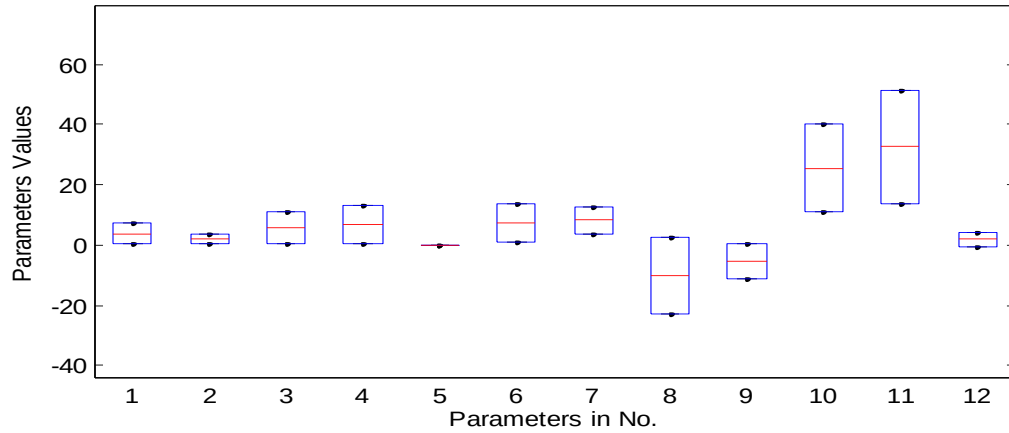


Figure 5.22: Spectrum of 'AUSTRALIA' word

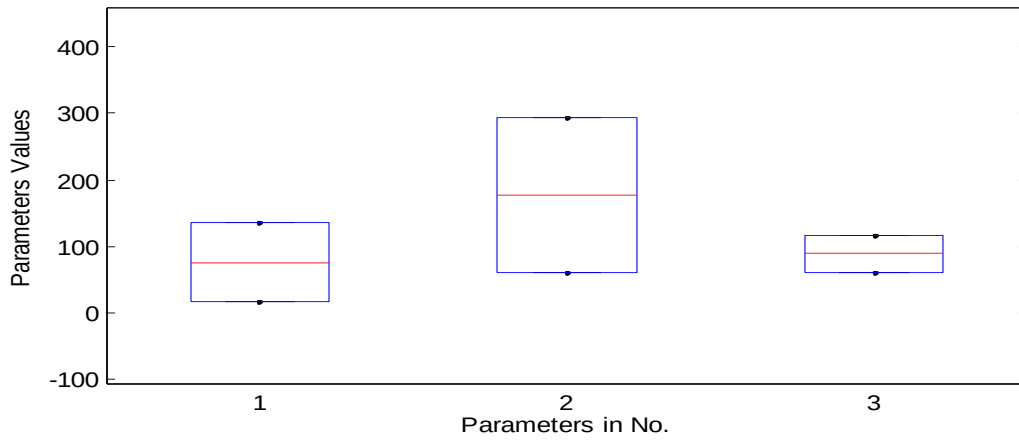
By Spectrum Analysis

Table 5.22(a): Parameters of word 'AUSTRALIA'

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	6.5942	7.2082	7.5463	7.3002	7.4196	7.21	0.37
MEDIAN	3.0222	3.2912	3.9089	3.5055	3.6943	3.48	0.34
STD	10.5449	10.852	10.9531	10.8734	11.4056	10.93	0.31
RMS	12.4367	13.0275	13.3006	13.0964	13.6063	13.09	0.43
MAX	130.4762	153.3329	134.1389	112.7429	147.9039	135.72	15.94
MIN	0.0133	0.0082	0.0101	0.0124	0.0063	0.010	0.003
SLOP	15.46	13.49	12.89	13.18	12.42	13.49	1.17
WIDTH	14	9	17	9	15	12.80	3.63
P.FREQ(Hz)	267.58	402.34	261.72	263.67	267.58	292.58	61.41
P. AMP(dB)	-23.01	-26.57	-20.14	-23.62	-20.22	-22.71	2.68
T. POW(dB)	-10.65	-10.52	-11.63	-11.3	-10.75	-10.97	0.47
THD(%)	47.61	50.13	24.92	45.59	32.42	40.13	10.92
THD+N (%)	118.44	220.18	68.33	107.01	72.16	117.22	61.50
IMD(%)	48.52	75.12	39.5	45.78	47.11	51.21	13.81
SNR(dB)	-1.47	-6.856	3.307	-0.58	2.83	-0.55	4.09



Graph 5.22(a): Insignificant Parameters from table 5.22(a)

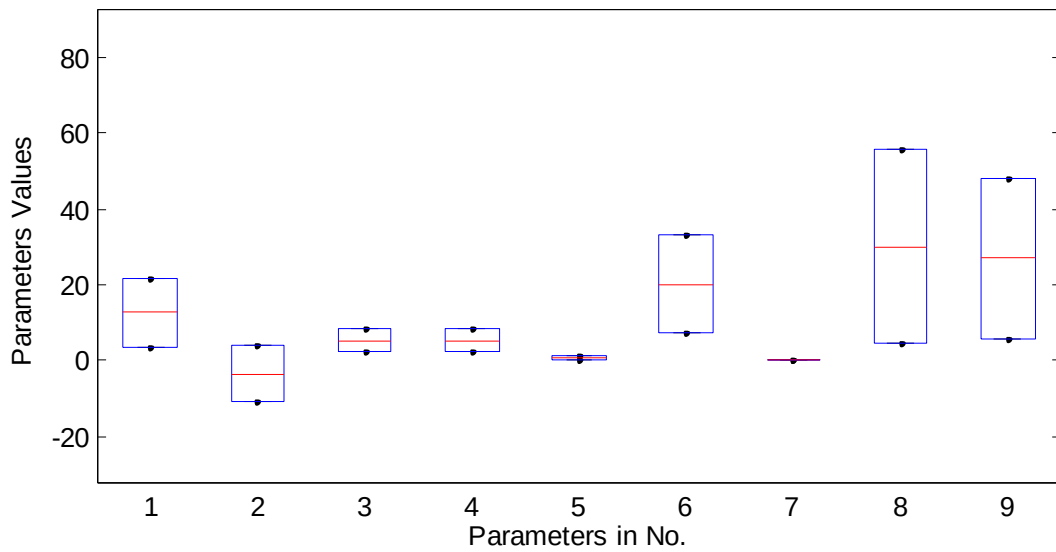


Graph 5.22(b): Significant Parameters from table 5.22(a)

By LPC Spectrum Analysis

Table 5.22(b): Parameters of word 'AUSTRALIA'

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	22.6568	15.5043	23.2947	25.2201	22.226	21.78	3.69
MIN	-14.071	-5.2753	-11.269	-14.846	-8.5274	-10.80	3.97
STD	9.6372	4.1226	8.9977	9.7701	9.4777	8.40	2.41
RMS	9.5995	4.1064	8.9626	9.7319	9.4406	8.37	2.40
SLOP	1.06	0.89	1.14	1.42	1.55	1.21	0.27
WIDTH	37	33	39	20	36	33.00	7.58
FRQ	0.0136	0.021	0.0136	0.0171	0.0136	0.016	0.003
MAXMAG	58	47.1	57.58	57.89	57.83	55.68	4.80
A.TION(dB)	46.3	42	49.1	47.3	57.2	48.38	5.58



Graph 5.22(c): Significant Parameters from table 5.22(b)

23 JAPAN

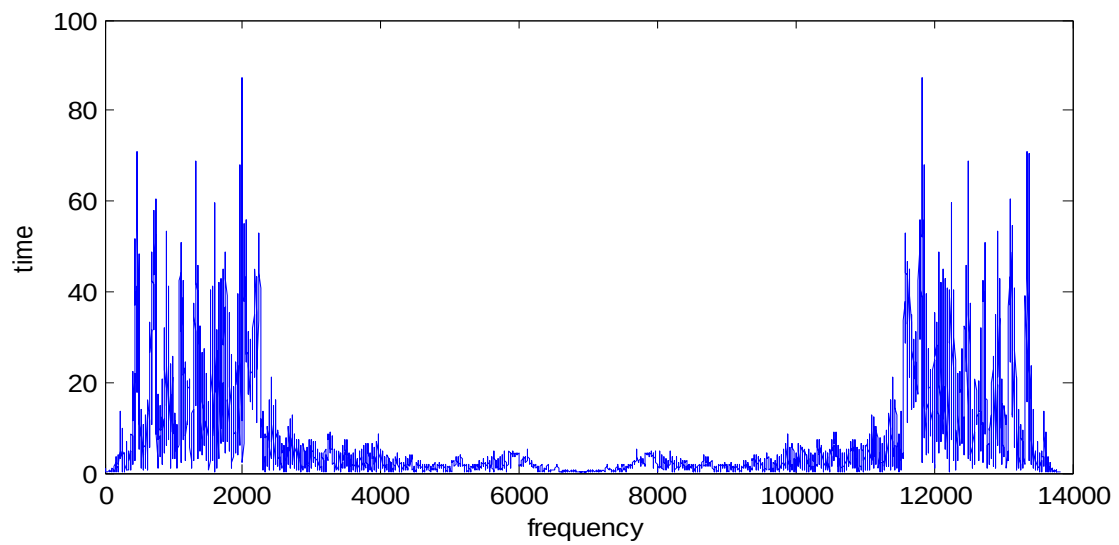
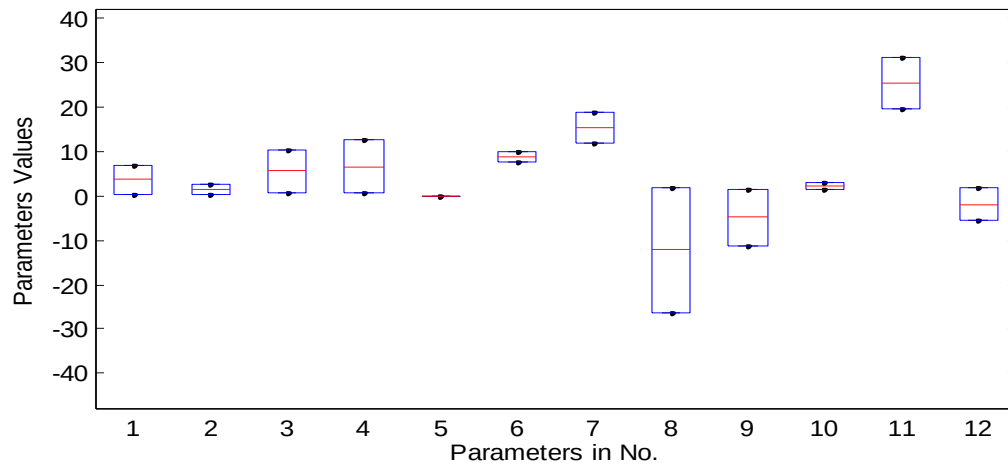


Figure 5.23: Spectrum of 'JAPAN' word

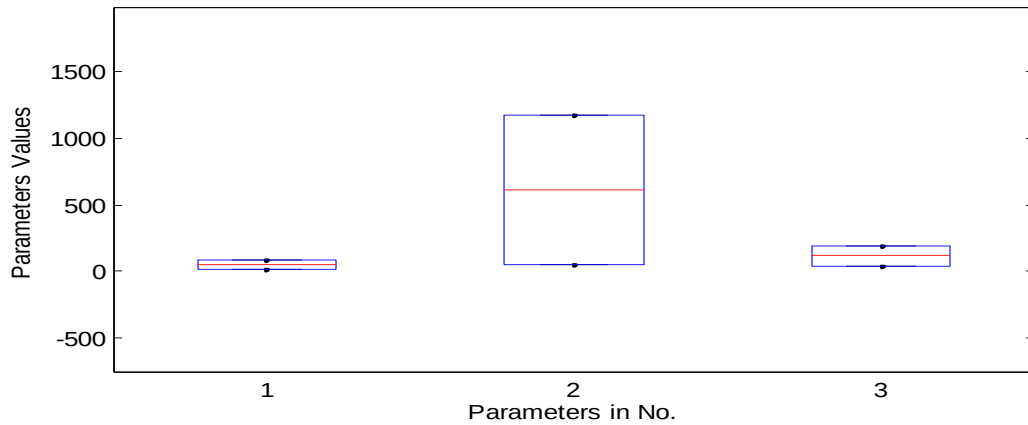
By Spectrum Analysis

Table 5.23(a): Parameters of word ‘JAPAN’

	S1	S2	S3	S4	S5	S6	Avg	Stdev
MEAN	6.3613	7.609	6.7819	6.7349	6.7279	6.4825	6.78	0.44
MEDIAN	2.4015	2.7445	2.5363	2.2246	2.3623	2.2272	2.42	0.20
STD	9.8534	11.3927	10.0581	11.2833	10.6027	9.8244	10.50	0.71
RMS	11.728	13.6996	12.1306	13.14	12.5568	11.77	12.50	0.79
MAX	58.9933	87.0887	82.772	103.6519	86.76	88.3081	84.60	14.47
MIN	0.0041	0.0066	0.0047	0.0057	0.0112	0.007	0.007	0.003
SLOP	18.7	6.958	2.8	6.46	20.45	3.72	9.85	7.72
WIDTH	4	21	31	23	4	29	18.67	11.94
P.FREQ(Hz)	1232.42	1150.39	1123.05	1161.11	1236.33	1101.56	1167.48	55.87
P.AMP(dB)	-28.07	-25.69	-27.85	-27.44	-26.59	-22.78	-26.40	1.98
T.POW(dB)	-10.44	-9.8	-12	-10.7	-14.03	-10.88	-11.31	1.51
THD(%)	1.53	3.01	5.66	2.85	1.915	3.19	3.03	1.45
THD+N (%)	259.93	187.57	183.66	220.43	165.16	143.03	193.30	41.52
IMD(%)	35.5	36.62	11.62	64.49	12.48	25.87	31.10	19.59
SNR(dB)	-8.297	-5.463	-5.28	-6.866	-4.35	-3.109	-5.56	1.83



Graph 5.23(a): Insignificant Parameters from table 5.23(a)

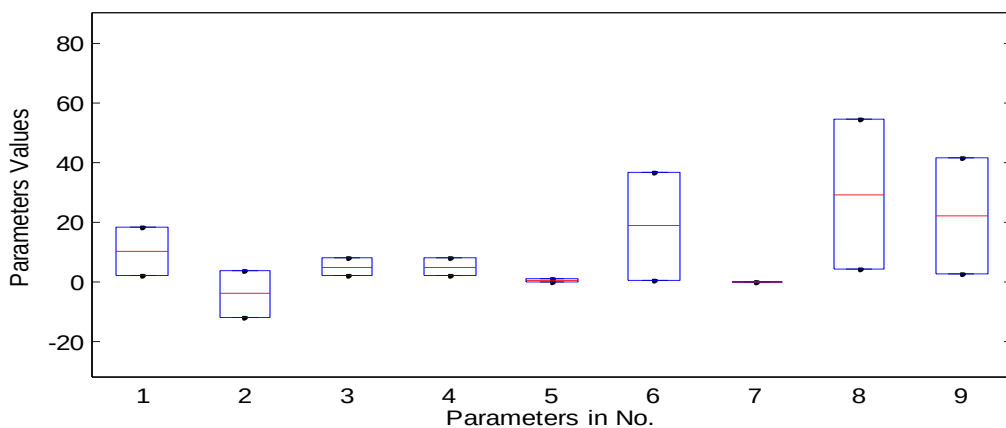


Graph 5.23(b): Significant Parameters from table 5.23(a)

By LPC Spectrum Analysis

Table 5.23(b): Parameters of word ‘JAPAN’

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	21.3317	17.5153	17.863	15.0189	20.378	18.42	2.50
MIN	-10.188	-16.357	-6.007	-14.028	-11.293	-11.57	3.93
STD	8.8575	10.375	4.5729	7.9858	8.649	8.09	2.15
RMS	8.823	10.375	4.5551	7.9548	8.6281	8.07	2.15
SLOP	0.77	0,80	1.4	0.7	0.75	0.95	0.39
WIDTH	37	36	36	38	36	36.60	0.89
FRQ	0.0117	0.0117	0.0156	0.0117	0.0149	0.013	0.002
MAXMAG	57.69	59.11	49.05	56.04	50.01	54.38	4.57
A.TION(dB)	43.8	39.5	43.7	37.5	42.5	41.40	2.79



Graph 5.23(c): Significant Parameters from table 5.23(b)

24 CHINA

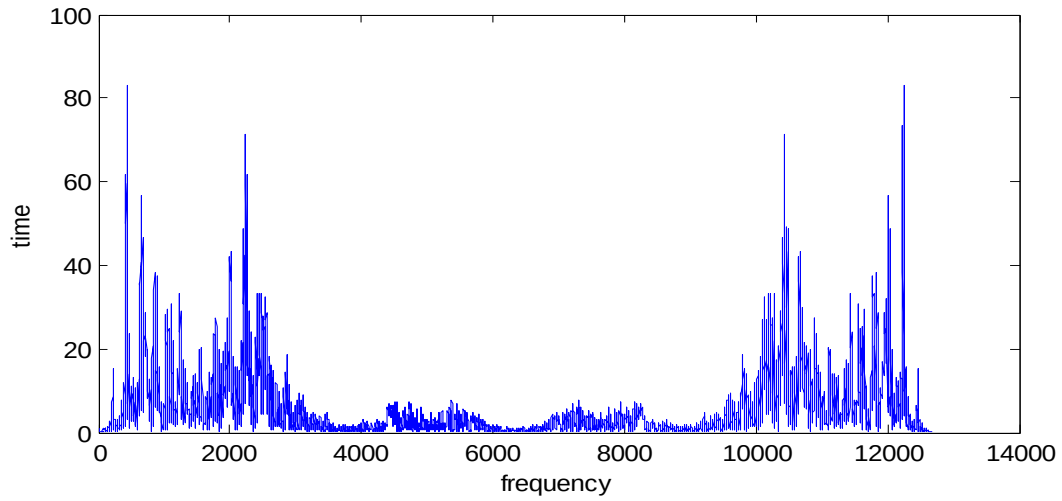
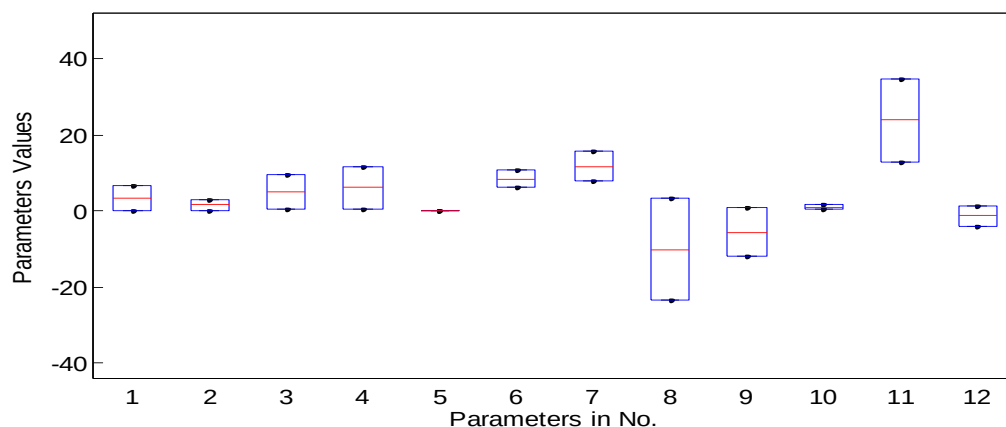


Figure 5.24: Spectrum of 'CHINA' word

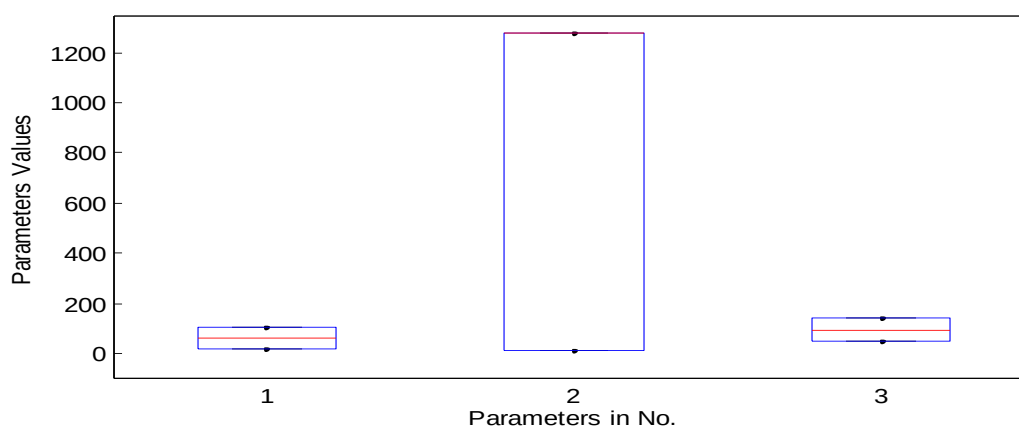
By Spectrum Analysis

Table 5.24(a): Parameters of word 'CHINA'

	S1	S2	S3	S4	S5	S6	Avg	Stdev
MEAN	6.9662	6.8345	6.29	6.7555	7.0285	6.79	6.78	0.26
MEDIAN	2.8889	3.485	2.9735	3.0755	3.2818	3.0329	3.12	0.22
STD	10.0847	8.9356	8.4719	9.723	10.0524	10.0288	9.55	0.68
RMS	12.2565	11.2494	10.5514	11.8391	12.2655	12.1109	11.71	0.68
MAX	134.4271	85.4927	83.1511	118.7074	116.0164	95.2075	105.50	20.63
MIN	0.0102	0.0104	0.0118	0.0161	0.0023	0.0056	0.009	0.005
SLOP	9.92	3.77	8.31	19.82	16.49	6.16	10.75	6.19
WIDTH	28	21	9	10	9	17	15.67	7.79
P.FREQ(Hz)	1300.78	1283.2	1267.58	1279.3	1273.44	1265.62	1278.32	12.88
P.AMP(dB)	-18.31	-23.17	-28.6	-23.64	-22.51	-23.21	-23.24	3.28
T.POW(dB)	-10.85	-11.98	-13.51	-12.23	-11.68	-11.15	-11.90	0.94
THD(%)	1.409	1.54	1.76	1.57	2.13	2.53	1.82	0.43
THD+N (%)	70.95	141.13	214.31	146.59	126.86	160.37	143.37	46.59
IMD(%)	25.13	28.61	17.41	39.19	48.6	49.49	34.74	13.11
SNR(dB)	-3.167	-2.99	-6.621	-3.32	-2.701	-4.103	-3.82	1.45



Graph 5.24(a): Insignificant Parameters from table 5.24(a)

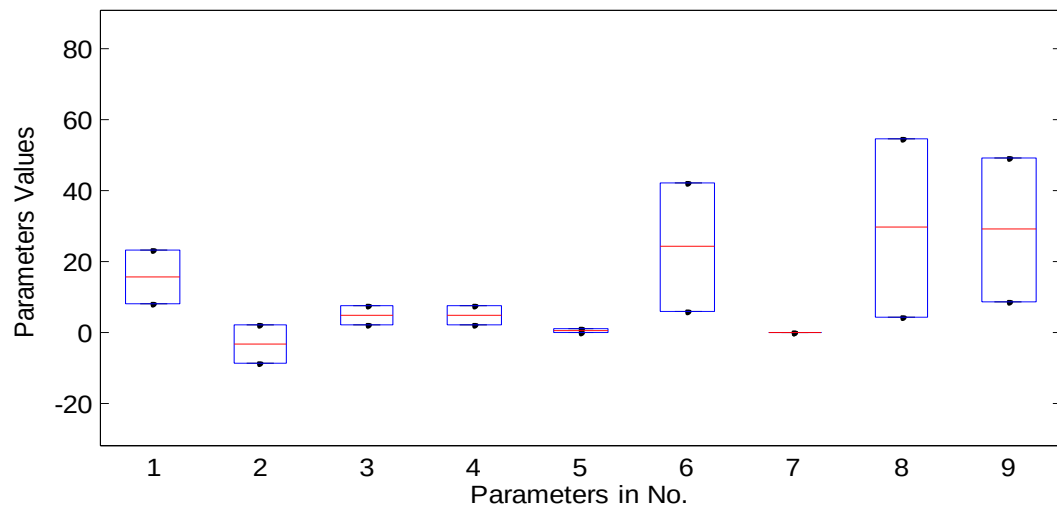


Graph 5.24(b): Significant Parameters from table 5.24(a)

By LPC Spectrum Analysis

Table 5.24(b): Parameters of word 'CHINA'

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	23.8222	30.9883	11.1357	20.4397	30.7729	23.43	8.24
MIN	-6.5392	-12.059	-6.8203	-9.3736	-7.9826	-8.55	2.26
STD	6.4684	10.1723	4.2553	8.6432	8.1842	7.54	2.26
RMS	6.4431	10.132	4.2387	8.6094	8.1522	7.52	2.25
SLOP	0.9	1.09	0.66	0.82	1.02	0.90	0.17
WIDTH	53	39	42	38	40	42.40	6.11
FRQ	0.0351	0.0117	0.0253	0.0117	0.0136	0.019	0.010
MAXMAG	52.65	58.94	47.79	57.75	55.82	54.59	4.48
A.TION(dB)	63.6	45.4	40.3	45.5	51.1	49.18	8.92



Graph 5.24(c): Significant Parameters from table 5.24(b)

25 RDX

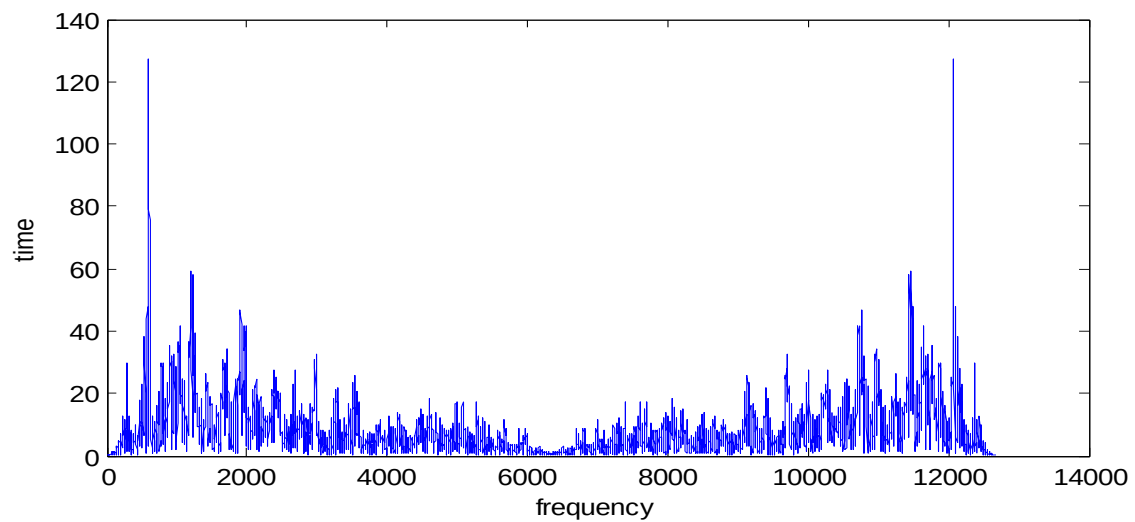
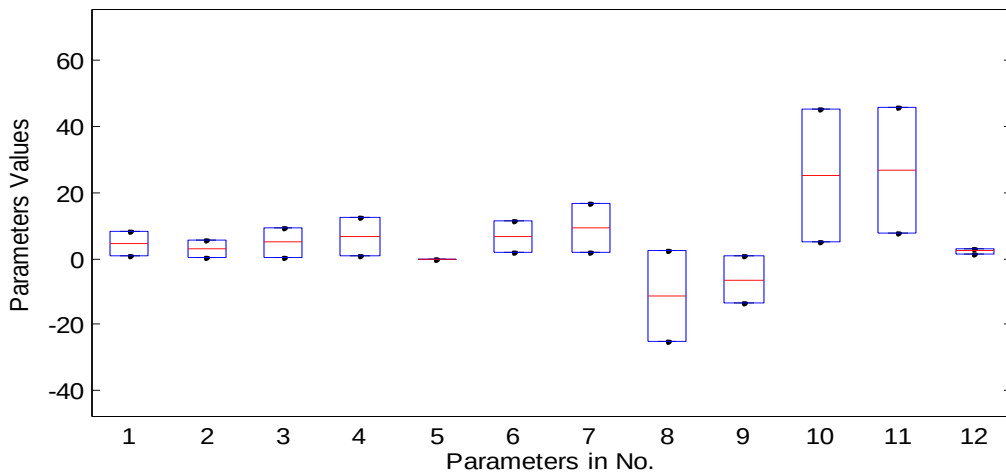


Figure 5.25: Spectrum of 'RDX' word

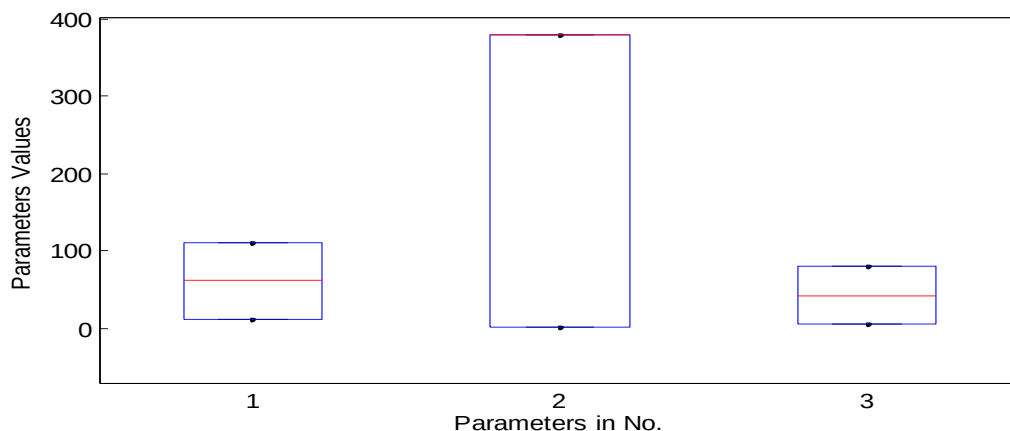
By Spectrum Analysis

Table 5.25(a): Parameters of word ‘RDX’

	S1	S2	S3	S4	S5	Avg	Stdev
MEAN	8.8561	8.3709	7.293	8.2931	7.471	8.06	0.66
MEDIAN	6.27	5.2272	5.3901	6.2934	5.3926	5.71	0.52
STD	9.1187	10.021	9.2931	9.1193	9.0127	9.31	0.41
RMS	12.7112	13.057	12.049	13.294	11.2936	12.48	0.81
MAX	127.0073	93.0456	112.834	109.538	115.843	111.65	12.30
MIN	0.0155	0.0061	0.0137	0.0141	0.0139	0.01	0.00
SLOP	13.75	9.25	12.49	10.92	12.09	11.70	1.70
WIDTH	15	20	16	17	16	16.80	1.92
P.FREQ(Hz)	376.45	378.91	380.86	379.59	377.47	378.66	1.74
P.AMP(dB)	-23.68	-29.18	-23.42	-25.84	-23.79	-25.18	2.43
T.POW(dB)	-13.63	-14.45	-13.91	-13.98	-12.47	-13.69	0.74
THD(%)	49.35	47.6	36.92	44.76	49.31	45.59	5.19
THD+N (%)	82.07	79.34	72.99	88.25	78.49	80.23	5.57
IMD(%)	36.82	48.1	54.74	50.48	38.97	45.82	7.66
SNR(dB)	2.52	-4.047	2.73	3.779	2.84	1.56	3.17



Graph 5.25(a): Insignificant Parameters from table 5.25(a)

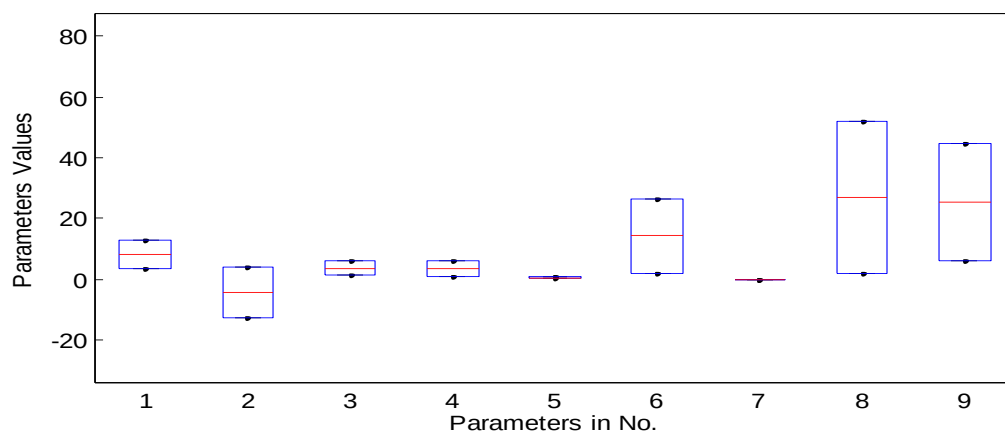


Graph 5.25(b): Significant Parameters from table 5.25(a)

By LPC Spectrum Analysis

Table 5.25(b): Parameters of word ‘RDX’

	S1	S2	S3	S4	S5	Avg	Stdev
MAX	16.974	11.324	8.451	14.492	13.294	12.91	3.22
MIN	-11.606	-17.121	-7.9461	-15.62	-10.283	-12.52	3.79
STD	6.398	7.16	4.3065	6.1934	5.318	5.88	1.10
RMS	6.373	7.132	4.2896	6.238	5.394	5.89	1.08
SLOP	0.82	0.87	0.7	0.83	0.73	0.79	0.07
WIDTH	29	25	25	28	26	26.60	1.82
FRQ	0.0156	0.0117	0.0117	0.0129	0.015	0.01	0.00
MAXMAG	53.12	54.06	49.12	52.36	50.37	51.81	2.02
A.TION(dB)	46.9	35.9	52.4	43.78	44.78	44.75	5.97

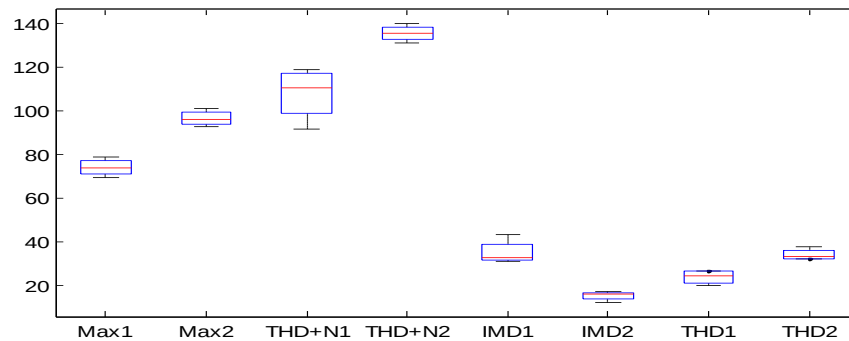


Graph 5.25(c): Significant Parameters from table 5.25(b)

5.2 DISCUSSION

From the results twenty four parameters of different words are obtained. The tables1-25 shows the parameter values. Each word has their own parameter value but when analysis the single word for five times then their parameter values changes. When these parameter values are found then we find that every parameter has their range in which that parameter lying. Every word parameters has their bounded values, out of these range values the parameter changes. From these values we plot their graphs, which tell us about their range values. These graphs clearly show the range of values in which every parameters lying. Some words their tons are same when they speaking like Boom - Bomb, But – Put, Kill – Fill, Singh – King but when we study these words then we get some parameters which differentiate these words to each others. These words have their own parameters which are same to each other's but at least some parameters are different which tells about the correct word.

FILL-KILL



Graph 5.26: Differentiate the word by some parameters

In this graph, odd no. is the parameter of Fill word and even no. is the parameter of Kill word. Max1 – 2 are Maximum Peak, THD+N1 – 2 are Total Harmonic Distortion+Noise, IMD1 - 2 are Inter Modulation Distortion, THD1 – 2 are Total Harmonic Distortion. Here same parameters of both words in which their values are very large difference which differentiate the correct word which are finding. One parameter value is very small but another word of same parameter value is very large which tell us about the word. Every parameter lying in bounded values and out of range this value the parameter has changed.

Fisher Linear Discriminant Analysis

$$LDA = \left| \frac{\mu_1 - \mu_2}{\sqrt{(\sigma_1^2 + \sigma_2^2)}} \right| \quad (1)$$

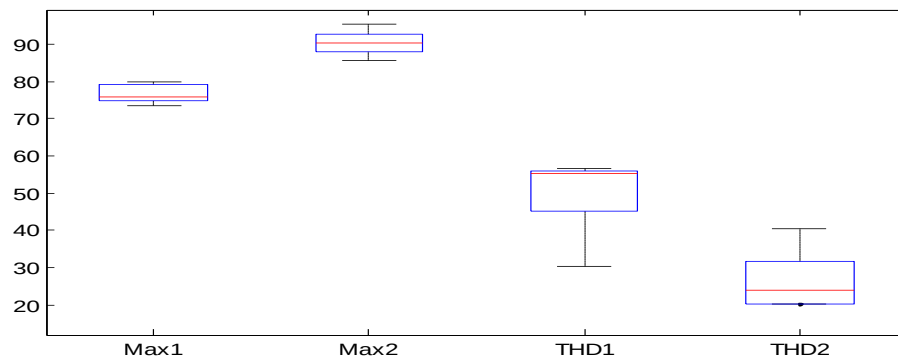
Where μ_1, μ_2 are the mean of parameters of words Fill and Kill. σ_1, σ_2 are the standard deviation of parameters of words Fill and Kill. By use this formula we Linear Discriminant Analysis of the same parameters of different words. [20] As shown below table 5.26.

Table 5.26: Parameters show their LDA for words ‘FILL-KILL’

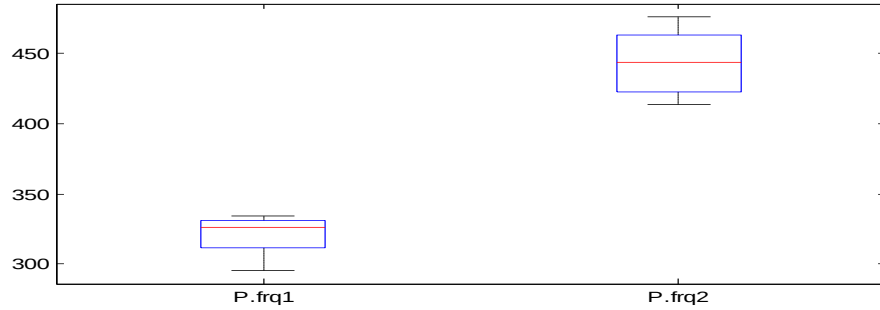
Parameters	LDA
Max _(FILL-KILL)	5.69
IMD _(FILL-KILL)	3.31
THD _(FILL-KILL)	2.55
THD+N _(FILL-KILL)	2.17

Higher the LDA values which more differentiate the same words. We show only upper range values (above 1), lower range values are insignificant for differentiate the correct word. Which parameter has high LDA value that parameter are more differentiate the words than lower LDA value. 1st Maximum parameter is differentiating the word, 2nd IMD, 3rd THD and then THD+N parameter.

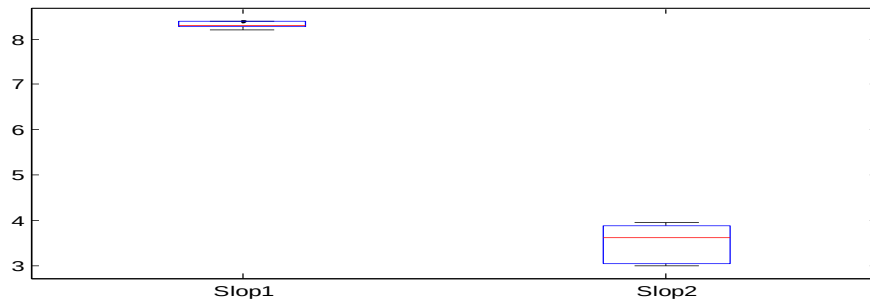
BOOM-BOMB



Graph 5.27(a): Differentiate the word by some parameters



Graph 5.27(b): Differentiate the word by P.frq parameter



Graph 5.27(c): Differentiate the word by Slop parameter

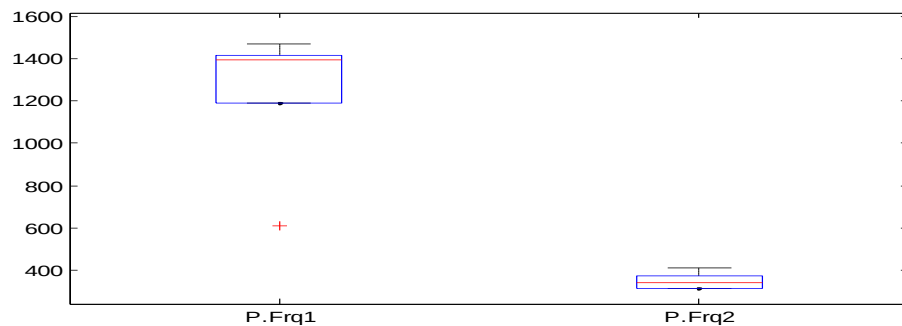
In these graphs, odds are Boom and even are Bomb parameters; they are same parameters of both words. In graph 5.27(a): Max1 – 2(Maximum Peak), THD1 – 2(Total Harmonic Distortion) are shown. In graph 5.27(b): P.Frq1 – 2(Peak Frequency) parameter is shown. In graph 5.27(c): Slop1 – 2(Slop of Peak) parameter is shown. Similarly here same parameters of both words in which their values are very large difference which differentiate the word. One parameter value is very small but another word of same parameter value is very large. Every parameter lying in bounded values and out of range these values the parameter has changed. Similarly, find LDA from above formula as shown below table 5.27.

Table 5.27: Parameters show their LDA for words ‘BOOM-BOMB’

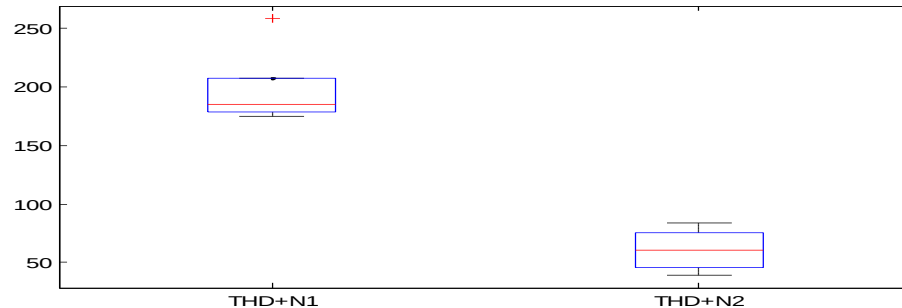
Parameters	LDA
Slop _(BOOM-BOMB)	10.54
P.Freq _(BOOM-BOMB)	4.17
Max _(BOOM-BOMB)	3.11
THD _(BOOM-BOMB)	1.63

Higher the LDA values which more differentiate the same words. We show only upper range values (above 1), lower range values are insignificant for differentiate the correct word. Which parameter has high LDA value that parameter are more differentiate the words than lower LDA value. 1st Slop of Peak parameter is differentiating the word, 2nd P.Freq, 3rd Maximum Peak and then THD parameter.

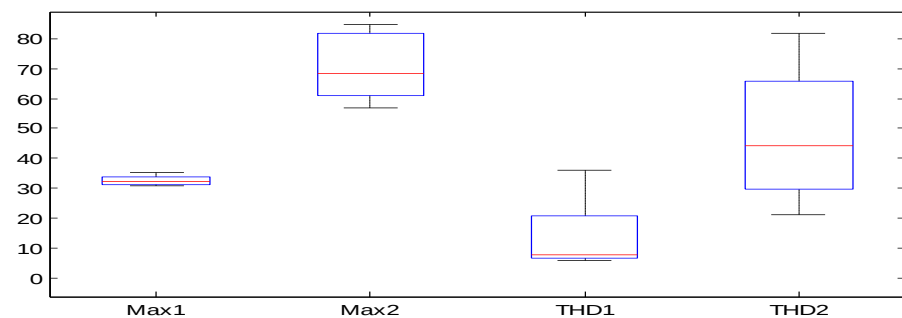
BUT-PUT



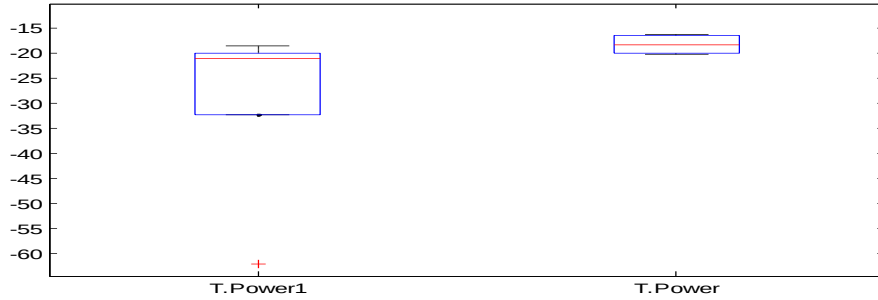
Graph 5.28(a): Differentiate the word by P.Frq parameter



Graph 5.28(b): Differentiate the word by THD+N parameter



Graph 5.28(c): Differentiate the word by some parameters



Graph 5.28(d): Differentiate the word by T. Power parameter

In these graphs, odd no. is the parameters of word ‘BUT’ and even no. is the parameters of word ‘PUT’. In graph 5.28(a): Peak Frequency, graph 5.28(b): THD+N, graph 5.28(c): Maximum Peak and THD parameter, graph 5.28(d): T. Power parameter are shown. Here same parameters of both words in which their values are very large difference which differentiate the word. Every parameter lying in bounded values and out of range this value the parameter has changed. Similarly we find LDA from above formula as shown below table 5.28.

Table 5.28: Parameters show their LDA for words ‘BUT-PUT’

Parameters	LDA
$\text{Max}_{(\text{BUT-PUT})}$	3.18
$\text{P.Freq}_{(\text{BUT-PUT})}$	2.49
$\text{THD}_{(\text{BUT-PUT})}$	1.16
$\text{THD+N}_{(\text{BUT-PUT})}$	0.69
$\text{T.Power}_{(\text{BUT-PUT})}$	0.57

Higher the LDA values which more differentiate the same words. We show only upper range values(above 1), lower range values are insignificant for differentiate the correct word. Which parameter has high LDA value that parameter are more differentiate the words than lower LDA value. 1st Maximum parameter is differentiate the words, 2nd P.Freq, 3rd THD, 4th THD+N and then T.Power parameter.

5.3 Accuracy

To find accuracy from this method we take 25 words and by analyzed find only 22 words, 3 words are not found, as shown in table 5.29.

Table 5.29 Identified words

Sr.No.	Words	Identification
1	‘ARMY’	Yes
2	‘POLICE’	Yes
3	‘BOOM’	Yes
4	‘BOMB’	Yes
5	‘BUS’	Yes
6	‘BUT’	Yes
7	‘PUT’	Yes
8	‘FILL’	Yes
9	‘KILL’	Yes
10	‘SINGH’	Yes
11	‘KING’	Yes
12	‘KITE’	Yes
13	‘NIGHT’	Yes
14	‘PUNJAB’	Yes
15	‘INDIA’	Yes
16	‘PAKSITAN’	Yes
17	‘MINISTER’	Yes
18	‘MURDER’	No
19	‘ENGLAND’	Yes
20	‘CANADA’	No
21	‘AMERICA’	Yes
22	‘AUSTRALIA’	No
23	‘JAPAN’	Yes
24	‘CHINA’	Yes
25	‘RDX’	Yes

Find the accuracy from this table:

$$\text{Accuracy result} = \frac{22}{25} \times 100 = 88$$

So percentile accuracy = 88 %

By use this method we find accuracy is 88 % which are very helpful for us to recognize the correct word.

CONCLUSION AND FUTURE WORK

Conclusion

In this work, 50 words are collected and analyzed on them, but only 25 words are shown. Each word has five samples and their spectrums parameter values are obtained using MATLAB. FFT spectrum and LPC coefficients methodologies are used to find these parameters. Twenty four parameters of single word are obtained. The result shows that these parameters vary between some values. Every word parameter has their bounded values in which that parameter varies. Each word has some specific range of these parameters. Some words are same but they still have some different parameters which tell us about the word. e.g in speech FILL and KILL words are similar and have some parameters which are same, but even though parameters like Maximum Peak, THD, THD+N, IMD are different which can be differentiate between them. So to recognition of any word, first of all values of the selected parameters are obtained. When these parameters values lies in between the bounded values then we find the correct word. By Fisher LDA and Box Plot methods are used to find those parameters likes Maximum Peak, THD, THD+N, IMD, Slop of Peak, Peak Frequency which differentiate the same words than all other twenty four parameters and find the correct word. The parameter with high LDA has maximum capability to differentiate the same words. If we take same word sample ten times collected from different persons, then their bounded values changes which makes the recognition task easier because every person has different notation, frequency, tone etc. Their range values are lying in large values or in large difference. The accuracy of this system developed is 88%. However this can be enhanced by using some AI techniques to train them.

We can use this method for recognizing the words which are spoken during on-line and off-line telephone conversations.

Future Work

A program can be created by using these significant parameters which have been obtained, capable of displaying the spoken word in textual form.

References

1. J. D. Tardelli, C. M. Walter, "Speech waveform analysis and recognition process based on non-Euclidean error minimization and matrix array processing techniques". *IEEE ICASSP*, pp. 1237-1240, 1986.
2. Takao Suzuki, Yasuo Shoji, "A new speech processing scheme for ATM switching systems". *IEEE, Digital Communications Laboratories, Oki Electric Industry Co. Ltd.*, Japan, pp. 1515-1519, 1989.
3. Hamid Sharkhzadeh, Li Deng, "Waveform based speech recognition using Hidden Filter Model parameter selection and sensitivity to power normalization". *IEEE Transactions on Audio and Speech Processing*, vol. 2, January 1994.
4. Andreas Spanias, Edward Painter, "An interactive GUI-Based tool for signal and speech processing". *IEEE FIE Conference*, pp. 95-100, 1998.
5. Mustafa Nazmi Kaynak, Qi Zhi, "Audio visual modeling for bimodal speech recognition". *IEEE Transactions on System, Man and Cybernetics*, vol. 34, pp. 564-570, July 2004.
6. Jose Leonardo Plaza Aguilar, David Báez López, "A Voice Recognition System for Speech Impaired People". *IEEE 14th International Conference on Electronics, Communications and Computers (CONIELECOMP)*, 2004.
7. Vikas C. Raykar, S. R. Mahadeva, "Speaker localization using extraction source information in speech". *IEEE Transactions on Audio and Speech Processing*, vol.13, September 2005.
8. Marcus Homberg, David Gelbart, "Automatic speech recognition with an adaptation model motivated by auditory processing". *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 43-49, January 2006.
9. Marco Grimadi, Fred Cummins, "Speaker identification using instantaneous frequencies". *IEEE Transactions on Audio, Speech and Language Processing*, vol.16, pp. 1097-1111, August 2008.
10. Ibrahim N. Abu-Isbeih, Khaled Dagrouq, Wael Ali-Sawalmeh, "Speaker identification wavelet transform based method". *IEEE 5th International Multi-Conference on Systems, Signals and Devices*, 2008.

11. Dr. E. Chandra, Mrs. C. Sunitha, "A review on speech and speaker authentication system using voice signal feature selection and extraction". *IEEE International Advance Computing Conference (IACC)*, pp. 1341-1346, March 2009.
12. Anup Kumar Paul, Dipankar Das, Md. Mustafa Kamal, "Bangla Speech Recognition System using LPC and ANN". *IEEE 7th International Conference on Advances in Pattern Recognition*, pp. 171-174, 2009.
13. Zufen Weng, Dongai Guo, Li.Lin, "Speaker recognition using weighted dynamic MFCC based on GMM". *IEEE*, pp. 285-288, 2010.
14. Ghai Shweta, S. Rohit, "Analyzing pitch robustness of PMVDR and MFCC features for children's speech recognition". *IEEE*, 2010.
15. Takahiro Murakami, Yoshihisa Ishida, "Fundamental frequency estimation of speech signals using MUSIC algorithm". *IEEE Acoust. Sci. & Tech.*, pp. 293-297, April 2001.
16. R. Manfred Schroeder, "Parameter Estimation in Speech: a Lesson in Unorthodoxy". *Processing of the IEEE*, vol. 58, pp. 707-712, May 1970.
17. M. Morales Perez, J. Echeverry Correa, "Feature extraction of speech signals in emotion identification". *30th Annual International IEEE EMBS Conference*, pp. 2590-2593, August 2008.
18. A. M. Abushariah, Teddy S. Gunawan, Othman O. Khalifa, "English Digits Speech Recognition System Based on Hidden Markov Models". *IEEE International Conference on Computer and Communication Engineering (ICCCE)*, May 2010.
19. Jan Korosi, Juraj Vojtko, Gregor Rozinaj, "Statistic Evaluation of Various Speech Parameters for Phonemes in Slovak Language". *IEEE 52nd International Symposium ELMAR*, pp. 375-378, September 2010.
20. R.A Fisher, G.T Mclachlan, "The Use of Multiple Measurement in Taxonomic Problems", (1936).
21. Chris Rowden, "Speech Processing" Mc. Graw. Hill, 1992.
22. Gorden E. Pelton, "Voice Processing" Mc. Graw. Hill, New York, 1993.