**SPEECH RECOGNITION IN THE PRESENCE**

**OF WIDEBAND NOISE**

A Major Project Report

Submitted in partial fulfillment of the requirements

for the award of the degree of

**MASTER of ENGINEERING**

in

**ELECTRONICS AND COMMUNICATION ENGINEERING**

Submitted by

**ASHISH KUMAR SAINI**

**University Roll No: 8514**

**College Roll No: 05/E&C/2K9**

Under the supervision and guidance of:

**PREM R. CHADHA**

Associate Profesor, Delhi College of Engineering



**DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING**

**DELHI COLLEGE OF ENGINEERING**

**UNIVERSITY OF DELHI, DELHI - 110042**

**2009-2011**

**DELHI COLLEGE OF ENGINEERING**

**(UNIVERSITY OF DELHI)**

**CERTIFICATE**

This is to certify that the Dissertation/Project report entitled **“SPEECH RECOGNITION IN THE PRESENCE OF WIDEBAND NOISE”** submitted by  **Mr.** **ASHISH KUMAR SAINI** (UNIVERSITY ROLL NO.- 8514 AND COLLEGE ROLL NO.- 05/E&C/2K9) in the requirement for the partial fulfillment for the award of the degree of **Master of Engineering** in **Electronics & Communication** at Delhi College of Engineering (University of Delhi) is a record of work carried out by him under my guidance and supervision in the academic year 2010-2011. To the best of my knowledge, the matter embodied in this thesis has not been submitted to any other University/ Institute for the award of any other degree or diploma.

Guided By:

**PREM R. CHADHA Dr. RAJIV KAPOOR**

Associate Professor Head of the Department

ECE Department ECE Department

Delhi College of Engineering Delhi College of Engineering

**ACKNOWLEDGEMENT**

It is a great pleasure to have the oppotunity to extend my heartiest felt gratitude to everybody who helped throughout the course of this project.

It is distinct pleasure to express my deep sense of gratitude and indebtedness to my learned supervisor Associate Prof. Prem R. Chadha, Delhi College of Engineering for his invaluable guidance, encouragement and patient reviews. His continuous inspiration, valuable suggestion,constructive criticism and active interest only made me to complete this dissertion. He kept on boosting me with time to put an extra ounce of effort to realize this work.

I am very thankful to **Dr. Rajiv Kapoor**, H.O.D, Electronics & communication, for his valuable suggestions and constant support.

I would also like to take this opportunity to present my sincere regards to all the faculty members of the Department for their support and encouragement.

I am grateful to my parents for their moral support all the time. They have been always around to cheer me up, in the odd times of this work.

I am also thankful to my class-mates for their unconditional support and motivation during this work.

**ASHISH KUMAR SAINI**

M.E. (Electronics & Communication)

University Roll No. 8514

College Roll No. 05/E&C/2K9

Department of Electronics & Communication

Delhi College of Engineering

**ABSTRACT**

This thesis describes the implementation of speech recognition algorithm under noisy environment. Reducing the level of background noise is very important in many communication systems. For example, most communication system are used in environments where the communication system needs to operate in the presence of high levels of car noise or street noise. Noise reduction also improves the performance of the speech recognition algorithms increasingly employed in a variety of real environments.

In this thesis we describe the implementation of its signal processing components, which provide a GUI-based environment to perform signal processing for speech recognition. This GUI-based configuration tool is presented in MATLAB. The tool described here deals with the Acoustic Front-end, which represent signal processing portions of a recognition system. More importantly, we present a thorough understanding and analysis for individual computational steps in the ASR algorithms.

**CONTENTS**

**Nomenclature Page Nos.**

**List of Figures**

**List of Tables**

**Abbreviations**

**Chapter- 1. INTRODUCTION** 1

**Chapter- 2. HISTORY OF ASR DEVELOPMENT 4**

**Chapter- 3. FUNDAMENTAL OF AUTOMATIC SPEECH**

**RECOGNITION 7**

3.1 How do humans do it? 8

3.2 How might computers do it? 10

3.3 Classification of ASR systems 11

3.3.1 Speaker-dependent versus independent system 11

3.3.2 Isolated versus continuous speech recognition

System 11

3.3.3 Small versus large vocabulary size 12

3.4 Automatic speech recognition process 13

3.4.1 Front-end analysis 14

3.4.2 Acoustic pattern recognition 15

3.4.3 Language analysis 16

3.5 Performance measurements of ASR 16

3.5.1 Recognition accuracy 16

3.5.2 Complexity 17

3.5.3 Robustness 17

3.6 Motivation and goal of this work 18

**CHAPTER 4. PROPOSED WORK USING SIGNAL PROCESSING**

**TECHNIQUES FOR FRONT-END 21**

4.1 Basic feature extraction principles 23

4.1.1 Pre-emphasis 23

4.1.2Frame blocking and windowing 25

4.1.3 Discrete Fourier Transform (DFT) computation 25

4.1.4 Spectral magnitudes 26

4.1.5 Mel-frequency filterbank 26

4.1.6 Logarithm of filter energies 29

4.1.7 Discrete Cosine Transformation (DCT) 29

4.1.8 Cepstral Weighting 30

4.1.9 Dynamic featuring 30

**CHAPTER 5. PRACTICAL PROBLEMS FACED AND SOLUTION IN ASR**

**APPLICATIONS 32**

5.1 Practical issues 33

5.2 Model of environment 35

5.3 End-point detection (EPD) 36

5.4 Spectral subtraction (SS) 38

**CHAPTER-6 HMM BASED ACOUSTIC MODELLING 40**

6.1 HMMs for ASR 41

6.2 HMM for the digit recognition task 45

6.3 Viterbi Search Engine 46

6.4 Viterbi trellis for “five” 47

6.5 Isolated word recognition (IWR) and Connected word

recognition (CWR) 48

6.5.1 Isolated word recognition (IWR) 48

6.5.2 Connected word recognition (CWR) 49

**CHAPTER-7 REMOVAL OF WIDEBAND NOISE FROM SPEECH SIGNAL**

**AND RESULT FOR SPEECH RECOGNITION 50**

7.1 Results for Noise Removal 52

7.2 Results for Speech Recognition 53

**CHAPTER-8 CONCLUSION AND FUTURE PROSPECTS 56**

8.1 Conclusion 57

8.2 Suggestions for future research 57

**REFERENCES** 59

**LIST OF FIGURES**

Figure 3.1: Voice interpretation by human ear. 9

Figure 3.2: Voice interpretation by Computers. 11

Figure 3.3 General block diagram of an automatic speech recognition

system. 14

Figure 4.1: Block diagram of ASR system. 22

Figure 4.2: Block diagram of MFCC front-end analysis. 24

Figure 4.3: Frequency response of mel-filter bank. 28

Figure 4.4: Mel scale of central frequency in each filter. 28

Figure 5.1 : Causes of acoustic variation in speech. 34

Figure 5.2 : Model of the environment. 36

Figure 5.3: End-point detection with boundary tb and te. 37

Figure 6.1 Example of Markov Model. 42

Figure 6.2 Markov Model sequence. 43

Figure 6.3 Example of Hidden Markov model. 44

Figure 6.4 Hidden Markov model sequence. 44

Figure 6.5 : HMM for digit recognition task. 45

Figure 6.6 : Viterbi trellis computation for the HMM. 46

Figure 6.7 : Viterbi trellis computation for “five”. 47

Figure 6.8: The Process of isolated word recognition. 48

Figure 6.9 : Search network of connected word recognition. 49

Figure 7.1 Removal of white band noise. 52

Figure 7.2 Speech signal “hello, this is matlab”. 52

Figure 7.3 Noise. 52

Figure 7.4 Noisy signal. 52

Figure 7.5 Denoised signal. 52

Figure 7.6 Three different utterance of word “one” after digitization. 53

Figure 7.7 Three different utterance of word “one” with noise. 53

Figure 7.8 Three different utterance of word “one” after removal of noise 53

Figure 7.9 Three different utterance of word “one” after framing. 53

Figure 7.10 Three different utterance of word “two” after digitization. 54

Figure 7.11 Three different utterance of word “two” with noise. 54

Figure 7.12 Three different utterance of word “two” after removal

of noise. 54

Figure 7.13 Three different utterance of word “two” after framing. 54

Figure 7.14 Three different utterance of word “three” after digitization. 55

Figure 7.15 Three different utterance of word “three” with noise. 55

Figure 7.16 Three different utterance of word “three” after removal

of noise. 55

Figure 7.17 Three different utterance of word “three” after framing. 55

**LIST OF TABLES**

Table 3.1: Classification of speech recognition difficulties. 13

Table 6.1 : Viterbi trellis computation for “five”. 47

**ABBREVIATIONS**

ASR : Automatic Speech Recognition.

DTW : Dynamic time warping.

HMM : Hidden Markov Models.

IWR : Isolated word recognition.

CWR : Connected word recognition.

MFCC : Mel-frequency cepstral coefficients.

LPC : Linear predictive coding.

LVCSR : Large Vocabulary Continuous Speech Recognition.