**CERTIFICATE**

This is to certify that the project entitled “**SPEAKER RECOGNITION**” submitted by Brij Mohan Purohit, Arjun Joshi, Ankur Bharti and Ashish Mathur to the Department of Computer Science Engineering, G. B. Pant Engineering College, Pauri Garhwal, in the partial fulfillment of the degree of **Bachelor of Technology in Computer Science Engineering**, during the year 2008-2012 is an authentic record of the work carried out under my supervision and guidance.

I wish them all success for their future endeavors.

**Supervisor**

**Mr. Ramesh Kumar**

**HOD CSED Assistant professor CSED**

**GBPEC Pauri GBPEC Pauri**

**PREFACE**

In this project, we propose to build a simple yet complete and representative automatic speaker recognition system, as applied to a voice based biometric system i.e. a voice based access control system. To achieve this, we have first made a comparative study of the MFCC approach with the Time domain approach for recognition by simulating both these techniques using MATLAB and analyzing the consistency of recognition using both the techniques.

Speech Recognition is the process of automatically recognizing a certain word spoken by a particular speaker based on individual information included in speech waves. This technique makes it possible to use the speaker’s voice to verify his/her identity and provide controlled access to services like voice based biometrics, database access services, voice based dialing, voice mail and remote access to computers.

Signal processing front end for extracting the feature set is an important stage in any speech recognition system. The optimum feature set is still not yet decided though the vast efforts of researchers. There are many types of features, which are derived differently and have good impact on the recognition rate. This project presents one of the techniques to extract the feature set from a speech signal, which can be used in speech recognition systems.

This Project Report is organized as follows:

In the chapter 1, the introduction to the project is presented.

In chapter 2, Introduction about basic speech signal.

In chapter 3, Introduction about automatic speaker recognition system is discussed.

In chapter 4, Feature extraction method is discussed.

In chapter 5, Algorithm used in this thesis is discussed.

In Chapter 6, GUI of all three methods will be given.

In chapter 7, Source code, conclusion, application of these project and Scope for future work will be discussed.

**ACKNOWLEDGEMENT**

Any accomplishment requires the serious efforts of many people and this work is no different. For the successful completion of this task, we express our gratefulness to our respected and learned guide **Mr. Ramesh Kumar** who stood by us patiently during every stage of this work.

We would like to say special thanks to our Head of the Department, **Mr. S.K. Verma** for giving us the valuable knowledge and suggestions whenever required during this work. We would like to thank to **Mr. Papender Kumar** and **Mr. S Dinkar**, faculty of Computer Science Engineering Department, for providing us the facilities and making us strive for the best possible output under their ardent observation.

In the end we would like to convey appreciation for our parents and friends who have been a constant source of inspiration and help till the successful completion of the venture.

Brij Mohan Purohit

Arjun Joshi

Ankur Bharti

Ashish Mathur

**Table of Contents**

|  |  |
| --- | --- |
| **Certificate** | **I** |
| **Preface** | **II** |
| **Acknowledgement** | **III** |
| **List of figures** | **VII** |
|  |  |
|  |  |

**Chapter 1 : Introduction**  **……………………………………………………………………1**

**1.1 Introduction……………………………………………………………1**

**1.2 Motivation……………………………………………………………...2**

**1.3 Objective……………………………………………………………….3**

**1.4 Outline of Project Report……………………………………………..3**

**Chapter2:Basic Acoustic and Speech Signal…………………….……..…………………..4**

* 1. **The Speech Signal …………………………………………………...4**
  2. **Speech Production……………………………………………...........5**
  3. **Properties of Human Voice………………………………………….6**

**Chapter 3 : Automatic Speech Recognition System………………………………………7**

**3.1 Introduction……………………………………………...………...……7**

3.1.1Speech Coding…………………………...……….….......7

3.1.2 Speech Synthesis………………........................................7

3.1.3 Voice Analysis………………………………………..…..8

3.1.4 Speech recognition………………………………............8

**3.2 Speech Recognition Basics…………………………………………….8**

3.2.1 Utterance………………….……………...……..………..8

3.2.2 Speaker Dependence……………………...……………...9

3.2.3 Vocabularies ……………………………...…...................9

3.2.4 Accuracy .......……..……..……………………………..9

3.2.5 Training……………………………..……………………..9

**3.3 Classification of ASR system ………………...………………………10**

3.3.1 Isolated Word………………..………………………….10

3.3.2 Connected Words……..……….…………..……………..10

3.3.3 Continuous Speech……………………………….............10

3.3.4 Spontaneous Speech……..………..……………………..10

3.3.5 Speaker Dependence……….……………………………10

**3.4 Why is Automatic Speaker Recognition Hard?.............................11**

3.4.1 Determining word boundaries…………...…………..11

3.4.2 Varying Accents……………….……..…...…………..11

3.4.3 Large Vocabularies………………………..………...11

3.4.4 Changing Room Acoustics…………………...………..11

3.4.5 Temporal Variance …………………………..………..11

**3.5 Speech Analyzer ……………………………………………………11**

3.5.1 Linear Predictive Coding………………..….…………12

3.5.2 Mel Frequency Cepstrum Coefficient..………..…….12

3.5.3 Perceptual Linear Prediction…………..........................12

**3.6 Speech Classifier ……………………………………….……………13**

3.6.1 Dynamic Time Wraping..…………..…..……………...13

3.6.2 Hidden Markov Model…………………...………….…13

3.6.3 Vecctor Quantization…………………………………..14

**Chapter 4 : Feature Extraction..…………………………………………..……………...15**

**4.1 Introduction ……………………………………………………….....15**

**4.2 Processing…….………..……………………………...………….........15**

**4.3 Features Extraction ……………………………….………………….15**

4.3.1 Frame Blocking……………….………………………..16

4.3.2 Windowing………………………...…………..……….16

4.3.3 Fast Fourier Transform…………….……………….......16

4.3.4 Mel Frequency Wraping .………...……………….…….16

4.3.5 Cepstrum………………………………………………..17

4.3.6 Mel Frequency Cepstrum Coefficient……..………..….18

**Chapter 5 : Algorithm ……………………….…………………….……………………...20**

**5.1 Introduction …………………..………….………………………....20**

**5.2 MFCC Approach……………….……….…………………….........20**

**5.3 Time Domain Analysis …………………..……………………….….24**

**5.4 Time Domain Approach………………………………………….......25**

**Chapter 6 : Sample Training & Recognition session………..…………………………....29**

**6.1 Main Menu …………………………...………………………….......29**

**6.2 GUI of MFCC method……………….…...…………………….......30**

**6.3 GUI of TDA …………………………………………………………..33**

**Chapter 7 : Source Code………………………………..……………………………….....35**

**7.1 MFCC Approach ……………………………………………….......35**

7.1.1 Training Phase………..…………………………………....35

7.1.2 Testing Phase……….………………..…………………....37

**7.2 Time Domain Approach……………………………..............................39**

7.2.1 Training Phase……………………………………………39

7.2.2 Testing Phase………………………..…………………..43

**Chapter 8 : Conclusion…………………………………….…………………………..…...48**

**8.1 Comparison ……………………..……………………………….…..48**

**8.2 Application…………………………..………………………….........50**

**8.3 Scope for future work ……………..…………………………….….51**

**List of Figures**

Fig. 1.1 Speaker Identification Training 2

Fig. 1.2 Speaker Identification Testing 2

Fig. 2.1 Schematic Diagram of the Speech Production/Perception Process 4

Fig. 2.2 Human Vocal Mechanism 5

Fig. 3.1 Utterance of "HELLO" 9

Fig. 3.2 Conceptual diagram illustrating vector quantization 14

Fig. 4.1 Feature Extraction Steps 15

Fig. 4.2 Filter Bank in Mel frequency scale 17

Fig. 4.3 Mel frequency scale 19

Fig. 5.2.1 MFCC Approach 20

Fig. 5.2.2 The word "Hello" taken for analysis 21

Fig. 5.2.3 The word "Hello" after silence detection 21

Fig. 5.2.4 The word "Hello" after windowing using Hamming window 22

Fig. 5.2.5 The word "Hello" after FFT 22

Fig. 5.2.6 The word "Hello" after Mel-warping 22

Fig 5.2.7 MFCC approach Algorithm 23

Fig 5.3.1 Overlapping windows in time domain 24

Fig 5.4.1 Plots of overlapping time domain hamming windows 26

Fig 5.4.2 Input speech "HELLO" 27

Fig 5.4.3 2500 speech samples extracted 27

Fig 5.4.4 Output of one of the overlapping windows 27

Fig 5.4.5 The frequency spectrum of the windowed signal showing the dominant

frequency 28

Fig. 6.1 Main Menu of the Speech Recognition Application 29

Fig. 6.2 GUI of MFCC method 30

Fig. 6.3 Testing Session GUI 31

Fig. 6.4 Access Granted 32

Fig. 6.3 Testing Session GUI 33

Fig. 6.6 GUI of Access Denied 34