Speaker Identification & Verification Using MFCC & SVM

Ahmed Sajjad[1], Ayesha Shirazi[2], Nagma Tabassum[3], Mohd Saquib[4], Naushad Sheikh[5]

[1]Professor, Dept. of Electronics & Telecommunication Engineering, Anjuman College of Engineering & Technology, Nagpur, Maharashtra, India

[2]Student of Graduation, Dept. of Electronics & Telecommunication Engineering, Anjuman College of Engineering & Technology, Nagpur, Maharashtra, India

[3] Student of Graduation, Dept. of Electronics & Telecommunication Engineering, Anjuman College of Engineering & Technology, Nagpur, Maharashtra, India

[4]Student of Graduation, Dept. of Electronics & Telecommunication Engineering, Anjuman College of Engineering & Technology, Nagpur, Maharashtra, India

[5]Student of Graduation, Dept. of Electronics & Telecommunication Engineering, Anjuman College of Engineering & Technology, Nagpur, Maharashtra, India

---***---

Abstract - *Speaker recognition is a developing source of security nowadays. Speaker recognition has a wide scope in future applications such as voice dialing, database access services, information services, security control, hospital, laboratories, industries etc. Speaker recognition is the process of automatically verifying and identifying the person who is speaking. This project is used to recognize the person who is speaking. Speaker recognition has two major parts speaker identification and speaker verification. The speaker recognition can be done by using two methods that is text dependent and text independent. This paper represents speaker identification and verification using speech dependent process. In this process first the features are extracted from voice and then those features are matched or verified in order to recognize the speaker. Here for feature extraction we are using MFCC (Mel Frequency Cepstral Coefficient) technique as it gives a great performance for making it robust, accurate, faster and computationally efficient. Also for feature matching SVM (Support Vector Machine) is used.*

Key Words*: *Speaker recognition, speaker identification, speaker verification, text dependent, text independent, feature extraction, MFCC, SVM

1. INTRODUCTION

The most convenient way of communication since ancient times is talking i.e. speaking to each other. Whenever we speak to someone we convey information in the form of words or voice or to be prudent say speech as this project is related to it. When air passes through the vocal tract of a person while speaking, inhaling etc. the vocal folds reflects this air which in turn produces speech. Hence speech is produced due to vibration in vocal system of a human body. Since every human being has a different vocal tract they produce a different sounds or speech. The aim of this project is to identify and hence verify different speeches or person. This recognition of a particular person through its speech automatically using a biometric device is done by using MFCC and SVM. MFCC and SVM are used as they give maximum accuracy as compared to LPCC, LPC, HPC, etc. The human pitch(an important characteristic of human voice) varies with the change in background noise(traffic, creeping of birds, unwanted sounds), human emotions(stress, happiness, envy), human health problems(cough, cold). These variations are easily eliminated in MFCC and SVM giving a high accuracy up to 95%. Also these are easily to work with.

2.1 Speech Production

Speech is produced with the help of vocal folds. The vocal system of a human being is responsible for the generation of speech . The human vocal system consist of nasal cavity, lips, teeth, glottis, tongue, palate, larynx, etc.

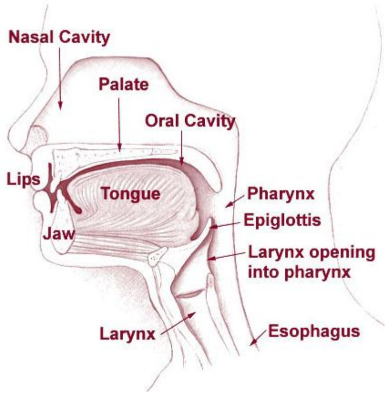

Figure 1: Human Speech Production System

“Speech is produced by air pressure waves emanating (emitting) from the mouth & the nostrils of a speaker.” As defined by Huang et al (2001)[1]. In other words speech is the ability to express feelings & thoughts by fluent sounds & gestures.

2.2 Speech Recognition

Speech recognition is nothing but conveying information to a computer, having it recognizes what we are saying, and finally doing this in real time.

Speech recognition has two functions identification and verification respectively. Speech identification is the process of identifying the speaker from the data base. It is a 1:N match. The voice given at the input is compared with voice available in data base until the voice is matched. If the voice is matches it means the speech is identified from N database otherwise the output as 'match not found'. Speech verification is the process of accepting or rejecting the identity of a speaker. It is a 1:1 match. This is a linear process where the input voice is checked with only one data and the result will be obtain as true or false, yes or no.

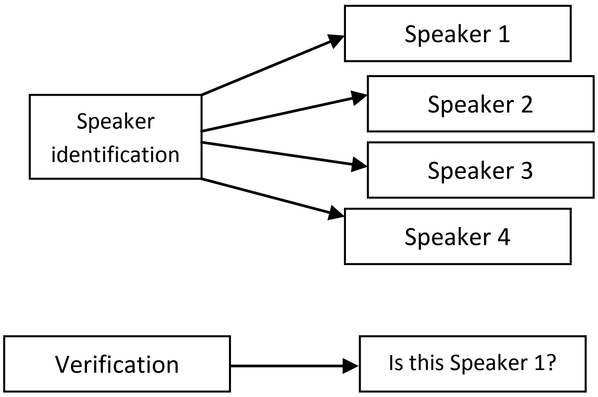

Figure 2: Speaker identification & verification process

Speech recognition can be done by using two processes:

Text Dependent: The text must be same at the time of feeding (preparing database) and while giving the input for recognition. This is know as text dependent process. In this process we can also use phrases or pins.

Text Independent: The process is said to be text independent, when the text at the time of feeding and verification is different. In this case there is no restriction over text

3. FEATURE EXTRACTION

It is the first and very important process of speaker recognition. It extracts the primary information from the speech and removes the other unnecessary data like background noise, other interruptions (stress, emotions, environmental conditions).

3.1 MFCC

Mel Frequency Cepstral Coefficient was introduced by Davis and Mermelstein in the 1980s. MFCC is most popular technique and commonly used in most of the application of speech signal for feature extraction[2]. We use MFCC because it is analogous to human hearing mechanism.
The MFCC consists of five major steps: pre-processing, windowing, FFT (Fast Fourier Transform), mel-frequency wrapping and cepstrum. The input signal is given to the MFCC and we get the desired coefficient known as MFCC.

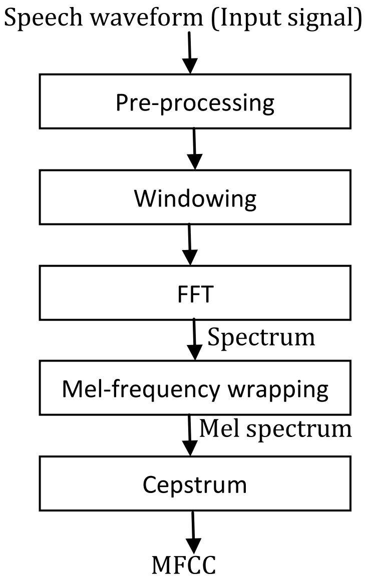

Figure 3: Extraction process of MFCC

Pre-processing: pre-processing includes filtering, filtering is converting the given voice signal in a form which is suitable for the computer. Pre-processing is segregating the voice part from the unvoiced part.

Windowing: It is used for minimizing the spectral distortion. For this we are using hamming window which is set to make frame blocking at 20-25 ms in order to achieve a stationary behavior. Hamming window provides continuity at the beginning and end of the each frame. It provides a better frequency resolution. The result of windowing is given as

$$Y(n) = X(n) \times w(n)$$

Where,
Y(n) – output signal
X(n) – input signal
w(n) – hamming window

FFT (Fast Fourier Transform): FFT is the most important step of MFCC is to construct the fast fourier transform of each frame which extract components from the signals at the rate of 10 ms. Fast fourier transform converts each N number of samples from time domain to frequency domain. The sizes of FFT are 512, 1024, 2048. It is used to obtain magnitude frequency response.

Mel-Frequency Wrapping: According to a psychological survey human presentation of frequency content of voice or speech is not proportional or can say does not follow a linear scale. For measurement of different pitch mel scale is used. "One mel is defined as one thousands of the pitch of a 1kHz tone[1]." Mel scale frequency can be approximated by equation:

$$B(f) = 2595 \log_{10} (1 + f / 700)$$

The simulation of spectrum is done by using filter bank. The triangular band pass frequency response is use as a filter bank. The position of filter bank is equally spaced by using mel-scale.

Cepstrum: The final step of MFCC is cepstrum in this step, Mel spectrum coefficients are converted into time domain by using DCT (Discrete Cosine Transform). The result will be obtained as MFCC.

4. FEATURE MATCHING

Feature matching is the process of identifying feature from two similar database. One knows as source and the other known as target.

4.1 SVM

SVM was developed by Vapinik in 1998. It is one of the most important developments in pattern recognition in the last 10 years[3]. As this technique gives more accuracy as compared to techniques like neural network, vector quantization etc.

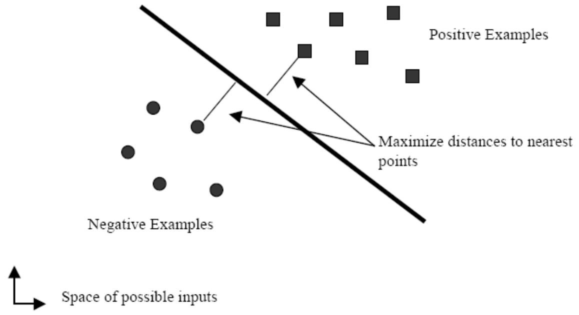

Figure 4: Linear Support Vector Machine

SVM is a simple and effective algorithm. It is a linear classifier[4] i.e. it can contain only two components at a time and gives a proportional output. Also it can be known as a comparator as it has a binary output it gives output as yes or no, accept or reject, 0 or 1 etc. In this project, we are using more than two components for better efficiency, hence we are using N number of SVM.

5. CONCLUSIONS

This paper describes a procedure for speaker recognition using MFCC and SVM. MFCC is used for feature extraction whereas SVM is used for feature verification. The importance of MFCC and SVM and why they are widely used is properly described in this paper. Instead of SVM techniques like GMM(Gaussian Mixture Model) and HMM(Hidden Markov Model) can be used in future as they are easier to use, require less data and gives better accuracy. The future application of this project are voice dialing in mobile phones and telephones, hands free dialing in Wireless Bluetooth headsets, biometric login to telephone aided shopping systems and numeric entry modules.

ACKNOWLEDGEMENT

I would like to thank Prof. Dr. Ahmed Sajjad Khan for the giving the idea of this project.

REFERENCES

[1] Huang, X, Acero, A. & Hon, H. "Spoken language processing – A guide to theory, algorithm, prentice hall PTR", New Jersey (2001).

[2] Jyoti B. Ramgire and Prof. Sumati M. Jagdale, "A survey on speaker recognition with various feature extraction and classification techniques", *IRJET*, Volume 3, Issue 4, April 2016, pp. 709-712.

[3] Geeta Nijhawan and M.K. Soni, "Speaker recognition using support vector machine", *International Journal of Computer Application,* Volume 87-No.2, February 2014.

[4] Simon Haykin, McMaster University, Hamilton, Ontario, Canada, Neural Networks a Comprehensive Foundation, 2nd edition, pp. 256-347.

BIOGRAPHIES

Prof. Dr. Ahmed Sajjad Khan
(PhD-Cellular Automata Modelling
and processing speech signal)

Ayesha Shirazi
Graduation Student
Nagpur University

Nagma Tabassum Shekh
Graduation Student
Nagpur University

Mohammad Saquib
Graduation Student
Nagpur University

Naushad Sheikh
Graduation Student
Nagpur University