BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

Course Number: EEE 312
Course Title: Digital Signal Processing Laboratory

Project Report

Submitted by:
Asif Billah 1706038
Tajwar Al Mamun 1706039
Habibullah Khan 1706040
Deb Indronil Sajib 1706041
Department: EEE
Level: 3, Term: 1
Section: A2
Group: 02

Contents

1 Abstract

Our goal of this project was to develop a MATLAB based system which can check attendance through identifying voice of a student.We have used mel frequency cepstral coefficients(MFCC) method to match a voice signal with database.The test data was tested against the train dataset.

In this project, an automatic speaker recognition system for attendance system is implemented by combining feature extraction and feature matching technique. Feature extraction method that is implemented by the Mel Frequency Cepstral Coefficients (MFCC). The distance between signals using dynamic time warping (DTW) is an algorithm that is used for the feature matching technique due to high accuracy but its simplicity.

2 Introduction

Voice can combine what people say and how they say it by two-facto
authentication in a single action. Other identifications like fingerprints, handwriting, iris, retina, face scans can also help in biometrics but voice identification is needed as an authentication that is both secure and unique. Voice can combine two factors, namely, personal voice recognition and telephone recognition. Voice recognition systems are cheap and easily understood by users. In today's smart world, voice recognition plays a very critical role in many aspects.Speech is one of the natural forms of human communication. Modern scientific technology has made it a security system based on a speaker recognition system. So the speaker recognition technology makes it possible to control access to secret services, for example, for giving commands to computers, phone access to banking, database services, shopping or voice mail and access to secure equipment by a speaker’s voice. Here we want to discuss a simplest model of speaker recognition system that could be applied to a speech of an unknown speaker but achieve more accuracy.Speaker Recognition is the process of automatically recognizing who is speaking on the basis of individual information included in speech signals.Another important application of speaker recognition technology is as a forensics tool.
Actually our main goal is to create a model using a speech recognition system where we will be able to take attendance through identifying the voice of a student.

3 Applications

In the last few years the use of biometric systems has become a reality. There are lots of commercial as well as personal applications where biometric is used for security purposes. Speaker verification has gained a huge acceptance in both government and financial sectors for secure authentication .Australian Government organization Centrelink, uses speaker verification for authentication of recipients using telephone transactions. Possible applications of speaker recognition are forensic investigation, telephone banking, access control, user authentication etc. . Speaker recognition has more potential than other biometrics such as face recognition, finger prints, and retina scans. The main advantage of speaker recognition over other biometrics is low cost, high acceptance and noninvasive character of speech acquisition. To develop a speaker recognition system, expensive equipment as well as direct participation of speakers is not required. Speaker recognition has the potential to eliminate the need of carrying debit card, credit card, remembering password for bank account or any other security locks and many other online services .With the continuous improvement in reliability of speaker recognition technology, its usability has increased. Nowadays, use of speaker recognition has become a commercial reality and part of consumer's everyday life. The performance of the speaker recognition system is vulnerable to change in speaker characteristics such as age, health problems, speaking environment etc.

4 Implementation

We have used the mel frequency cepstral coefficients(MFCC) method to match a voice signal with a database. The purpose is to extract features. Speech signal vector is taken as an input parameter

Speech works as a parameter for input .Few more parameters are taken as inputs.They are as follows.
Fs is the sampling frequency in (Hz).
TW is the analysis frame duration in (ms).
TS is the analysis frame shift in (ms).
ALPHA is the preemphasis coefficient.
WINDOW is a analysis window function
R is the frequency range (Hz) for filterbank analysis.
M is the number of filterbank channels.
N is the number of cepstral coefficients.
L is the liftering parameter.

mfcc_c function returns mel frequency cepstral coefficients as a vector MFCC
FBE is a matrix of filterbank energies.
FRAMES is a matrix of windowed frames(one frame per column).

Three things to be noted shortly:
(i)pre emphasis:using coefficient of ALPHA and speech and then doing first order FIR filter,output speech is then passed through butterworth filter.for this range i used and filtering results in new speech.
(ii) here firstly we do windowing and then framing.
(iii)Using “frames” and “nfft” we compute the magnitude spectrum “MAG”.

(MFCCs) computed from speech signal given in vector S and sampled at FS (Hz). The speech signal is first pre-emphasized using a first order FIR filter with pre-emphasis coefficient ALPHA. The pre emphasis speech signal is subjected to the short-time Fourier transform analysis with frame durations of TW (ms), frame shifts of TS (ms) and analysis window function given as a function handle in WINDOW. This is followed by magnitude spectrum computation followed by filterbank design with M triangular filters uniformly spaced on the mel scale between lower and upper frequency limits given in R (Hz). The filterbank is applied to the magnitude spectrum values to produce filterbank energies (FBEs) (M per frame). Log-compressed FBEs are then decorrelated using the discrete cosine transform to produce cepstral coefficients. Final step applies sinusoidal lifters to produce liftered MFCCs that closely match those produced by HTK.

5 Diagram

Test and train data > MFCC > Comparison between MFCC with DTW > Feature matching for the closest one

6 Explanation

When in the main code,k=1(in the loop) is activated then we are taking ‘new entry’ for ‘entry’, ‘id’,and ‘name’. Then the train function is called and train data are taken on run time.In addition we can take train data set off run time.

Here we take the test data set for only on run time. Then MFCC of test data samples are calculated.After that MFCC of train samples are calculated. In this step we compare and measure the distance for test and all train entries under one folder.Then we calculate average distance for that folder.Choosing the folder with minimum distance and return the name.

Finally we compare for ‘entry’,’id’,and ‘name’.When output is matched then result is ‘present’(access granted).If output is mismatched, the result will be ‘not present’(access denied).

7 Workflow

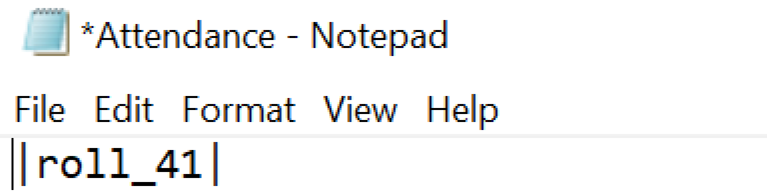
1. Checked various methods of voice recognition from various papers.
2. Get an idea about the parameters related to speech recognition.
3. Creating a database for different people.
4. Matlab implementation of function for MFCC.
5. Comparison between test values and database values.
6. Displaying the result.

8 Test Results

<u>When data of roll_41 are entered:</u>

Say Entry
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Say name
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Say Id
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Access Granted
Welcome roll_41>>

In the attendance sheet :



When data of Asif are entered:

Say Entry
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Say name
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Say Id
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Access Granted
Welcome Asif

When data of Asif with fan are entered or different environment:

Say Entry
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Say name
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Say Id
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Access Denied

When data of Nafis are entered:

Say Entry
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Say name
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Say Id
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Access Denied

When data of both Asif and Nafis are entered:

Entry and roll: Asif
Name: Nafis

Say Entry
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Say name
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Say Id
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Access denied

When data of unknown person are entered:

Say Entry
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Say name
Start Speaking
3
2
1
NOW!!!
Stop Speaking
Say Id
Start Speaking
3
2
1
NOW!!!

Stop Speaking
Access denied

9 Drawbacks

We can see from the previous section, the test results are not so accurate. The drawbacks are as follows:

1. Same environment and also the same device are required.

2. Keywords must be followed for both train and test data sets.

3. Sometimes access is denied but two or more tries should do it.

4. If any person is not enlisted in train data then his/her test data is matched with one of the enlisted one but it is highly unlikely that for enlisted ones the data is matched with the wrong person.

10 Conclusion and Future Scope

This report provides concise definition and discussion about the attendance system using speaker recognition technology. We have proposed an automatic speaker recognition system with the combined Mel Frequency Cepstral Coefficients (MFCC) that is used as feature extraction. Basic concepts of automatic speaker recognition systems, clustering techniques etc. have been discussed. Speaker recognition is a method of designing a system for the identity of an individual through their voices. Speaker recognition has a significant perspective as it is an appropriate biometric technique for security. The speaker recognition task is normally achieved by acquiring speech signal, feature extraction, pattern matching and obtaining match score.

This is the initial work about the attendance system using the speaker recognition task for us as this study is still ongoing. So we have a plan for this work, such as recognizing the speaker if the speech deals with crosstalk, laughter and uncharacteristic sounds. We want to proceed with machine learning for this type of work.

11 List of References

[1]Abdullah-al-MAMUN, M. D., Firoz Mahmud, Aminul Islam, and Syed Tauhid Zuhori. "Automatic Speaker Recognition system using Mel Frequency Cepstral Coefficients (MFCC) and Vector Quantization (VQ) approach."
[2] Automatic Speech Recognition using correlation analysis By Rajorshee Raha, Amab Pramanik
[3] An Enhanced Speech Recognition System By Suma Shankaranand, Manasa S, Mani Sharma, Nithya A.S, Roopa K.S., K.V. Ramakrishnan, International Journal of Recent Development in Engineering and Technology,Volume 2, Issue 3, March 2014.
[4] Mahdi Shaneh and Azizollah Taheri, ‖Voice Command Recognition System based on MFCC and VQ Algorithms‖, World Academy of Science, Engineering and Technology Journal, 2009.
[5] Nikolai Shokhirev, ‖Hidden Markov Models —, 2010.
[6] SPEECH RECOGNITION USING MATLAB By ASEEM SAXENA,AMIT KUMAR SINHA,SHASHANK CHAKRAWARTI,SURABHI CHARU, International Journal of Advances In Computer Science and Cloud Computing, ISSN: 2321-4058 Volume- 1, Issue- 2, Nov-2013.

THE END