EL – 413 Digital Signal Processing

Semester Project

VOICE RECOGNITION

Student Name: Tuba Siddiqui

Roll Number: 17B-042-CE Section: E

Maximum Marks	Simulations= 05	Project Presentation = 05	Project Report Evaluation= 05	Total = 15
Marks Obtained				
Remarks (if any)				

Experiment evaluated by

Instructor Name: Engr. Syeda Aimen Naseem/ Engr. Mehr un Nisa

Signature:

Table of Contents

Abstract

Voice Recognition system is a method of analyzing the input voice of the person with the help of its features. It than compares it with the features saved in the database for prerecorded signals. It displays an output that tells if any other audio of the same person is present in the database or not. If yes, it displays the name or the file number. Unauthorized entries can be a danger, to keep them in control and checking them frequently can reduce the chances of loss and problems. It is an effective method of securing things and places like lockers, phones, offices, etc. This is beneficial for measuring how many times any person tried to access the system. This system is currently limited to checking if already a record with same features exists or not.

I. Introduction

Technology of voice recognition can be roughly divided into two sub-areas, which are speech recognition and speaker recognition [1]. Speaker recognition is a method of analyzing the person who has sent its audio in the input by comparing it with the records in its database. Speaker identity is verified on the basis of its voice's features. Either pitch, speed, power energies, etc. Mainly it is used in security purposes to either give aces to or deny the person who has sent its audio in the input. In this system, the program would display the file number of the person whose audio is sent in the input. For signals, this is typically done via the cross-correlation function (of two signals) which is very similar to convolution. As such, it can be mathematically done via the FFT, which is specifically designed to be efficient [2].
The Discrete Fourier Transform (DFT) is a kind of FFT. The difference is DFT take longer time whereas FFT takes lesser time. DFT converts the signal from time domain to frequency domain, while FFT is just an implementation of DFT. FFT is far more efficient method of conversion. It transforms one function into another, which is called the frequency domain representation, or simply the DFT, of the original function (which is often a function in the time domain). But the DFT requires an input function that is discrete and whose non-zero values have a limited (finite) duration. Such inputs are often created by sampling a continuous function, like a person's voice [2]. This system will basically check whether any other audio of the tested person exist in the system already or not.

II. Literature Review

These papers discussed speaker recognition system, and apply it to a speech of an unknown speaker. Over the years, a lot of contribution has be done in this field. Due to be efficient working and providing safety, it is a demand at various places like banks, offices, army etc. Many methods have been use for this, like convolution neural network [2], Euclidean distance and humming method [3]. For best results, MFCC is majorly used. [5]

III. Methodology

a. Procedure

i. Background:

a. Input:

Figure 1

b. Output:

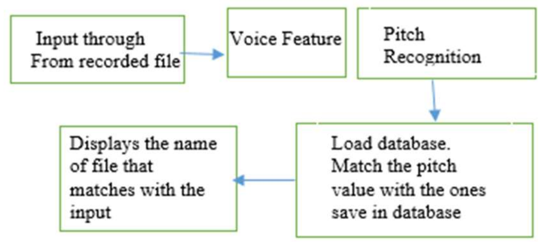

Figure 2

c. Method:

There are 3 major functions in this system.

- Voice Training,
- Voice Testing,
- Voice Feature.

i. **Voice Training:**
Already recorded samples of users are saved in the directory. In this function, system would call the function Voice Feature and save the pitch of each audio in the database.

ii. Voice Testing:
An input would be placed by typing in the name of file we want to test. It will call the Voice Feature function and match its pitch with the pitches in database and show the name of the files that matches the audio.

iii. Voice Feature:
This function calculates FFT of the audio and takes its real part as an output. Then it take the maximum
value from the frequencies came in FFT.

b. Source code

i. Voice Training:

```
clear all;
close all;
clc;
%% Store Features
for i=1:13

filename=strcat('E:\6thSem\dsp\VoiceRecog
nition\train\a',num2str(i),'.wav');
    b = audioread(filename);

    FE=VoiceFeatures(b);
try
    load database
    F=[F;FE];
    FN=[FN;i];
    database=[database;F;FN];
    save database.mat database F FN
catch
     F=FE;%F=features
     FN=i;
    save database F FN
end
end
```

ii. Voice Testing:

```
clear all;
close all;
clc;
%% Input for testing
name=input('Whose file do you want to
run? Enter name: ','s')
file=strcat('E:\6thSem\dsp\VoiceRecogniti
on\test\',name,'.wav');
b=audioread(file);
%% Feature Extraction
FE=VoiceFeatures(b);
%% Classify
load database
D=[];
for(i=1:size(F,1))%returns the length of
1st dimension of F
    d=sum(abs(F(i)-FE));
    D=[D d];
end
%% Smallest distance
sm=inf;
ind=-1;
for(i=1:length(D))
    if(D(i)<sm)
        sm=D(i);
        ind=i;
    end
end
%% Output
file_number=FN(ind);
sc=strcat('The file number of ', name, '
in training is: ');
disp(sc)
file_number
%% Plotting
%test file
[t,x]=audioread(file);
subplot(2,1,1)
plot(abs(t(:,1)))
xlabel('Time')
ylabel('Amplitude')
title('TEST')
%train file
subplot(2,1,2)
filet=strcat('E:\6thSem\dsp\VoiceRecognit
ion\train\a',num2str(file_number),
'.wav');
[v,s]=audioread(filet);
plot(abs(v(:,1)))
xlabel('Time')
ylabel('Amplitude')
title('TRAIN')
%Pitch
figure(2);
plot(real(fft(v)))
title('PITCH')
```

iii. Voice Feature:

```
function [xPitch]=VoiceFeatures(b)
F=fft(b(:,1));
%plot(real(F));
m=max(real (F));
xPitch=find(real(F) == m,1) %find out
first istance of maxima
```

VI. Results

For demonstration, I first ran the code of voice training. Firstly, all the saved files in .wav format are loaded one by one and the function calls voice feature function which gave pitch value for every audio file in directory. Lastly, in voice testing, an input is sent from the directory. I typed in ‘tuba.wav’ as name in input. The function matches the pitch of ‘tuba.wav’ with pitches in database and giving a file number in output, telling with which file the pitch of the audio matched.

```
Whose file do you want to run? Enter name: tuba

name =

tuba


xPitch =

   391

The file number oftuba in training is:

file_number =

    12
```

To match whether the tested and output file that program said is matched with, graph is shown:

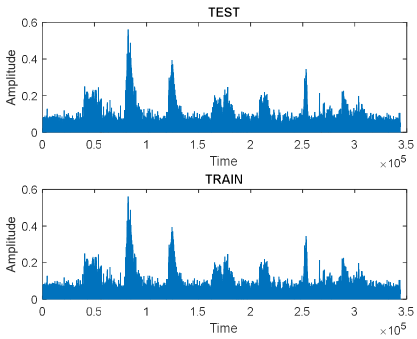

Figure 3

FFT:

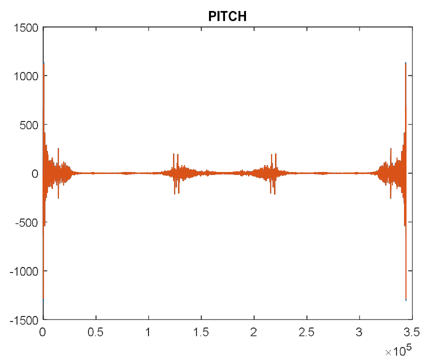

Figure 4

IV. Conclusion

This method is efficient for matching the audios in the input with already saved files to check which file is already saved with the same features.

V. Future work

In future, this can be expanded as a system that allows or deny the entry of the person to ensure security. In addition, an alarm system can be cascaded which would be triggered on false entry. It rather notifies by sending a message to the owner or an alarm voice can be generated.

VI. Reference

[1] L. Rabiner L, B. H. Juang, *Fundamentals of Speech Recognition*, Englewood Cliffs, NJ: Prentice - Hall International, 1993.

[2] Supaporn Bunrit, Thuttaphol Inkian, Nittaya Kerdprasop, and Kittisak Kerdprasop *Text-Independent Speaker Identification Using Deep Learning Model of Convolution Neural Network* International Journal of Machine Learning and Computing, Vol. 9, No. 2, April 2019

[3] *Speech Recognition Using Euclidean Distance* Akanksha Singh Thakur1 , Namrata Sahayam2 1ME IV Sem, 2Asst Professor

[4] *Implementation of a Voice-Based Biometric System*, ADARSH K.P., A. R. DEEPAK, DIWAKAR R, KARTHIK R. 2006-2007

[5] *Voice identification and recognition system* SOHAIB TALLAT, FARHAN SHAHID, ABDUL SAMAD , MATTI ULLAH ABBASI *DECEMBER 22, 2014*