Final Project Report

ITE1014-Human Computer Interaction

Professor Rajesh Kaluri

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

PROJECT TITLE

Speech Recognition using Correlation in MATLAB

GROUP MEMBERS

16BIT0251 – Vaibhav Verma

16BIT0013 – Kunal Pundir

INTRODUCTION

Speech is one of the most important medium by which a communication can take place. With the invention and widespread use of mobiles, telephones, data storage devices etc. has provided a major help in setting up of speech communication and its analyzing. The term and the basic concept of speech identification was began in the early 1960`s with exploration into voiceprint analysis which was somewhat similar to fingerprint concept.

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.

Speech recognition is the inter-disciplinary sub-field of computational linguistics that develops methodologies and technologies that enables the recognition and translation of spoken language into text by computers. It is also known as automatic speech recognition (ASR).

The system analyzes the person's specific voice and uses it to fine-tune the recognition of that person's speech, resulting in increased accuracy.

LITERATURE REVIEW

Proposed System

The structure of proposed framework comprises of two modules to be specific, Speaker Identification taken after by Speech Recognition.

Speaker Identification

Feature extraction is a procedure that concentrates information from the voice flag that is one of a kind for every speaker. Mel Frequency Cepstral Coefficient (MFCC) method is regularly used to make the unique finger impression of the sound documents. The MFCC depend on the known variety of the human ear's basic data transfer capacity frequencies with channels separated straightly at low frequencies and logarithmically at high frequencies used to catch the vital qualities of discourse.

These extricated components are Vector quantized utilizing Vector Quantization calculation. Vector Quantization (VQ) is utilized for highlight extraction in both the preparation and testing stages. It is a to a great degree effective portrayal of unearthly data in the discourse motion by mapping the vectors from vast vector space to a limited number of districts in the space called clusters. After component extraction, highlight coordinating includes the genuine technique to recognize the obscure speaker by contrasting separated elements and the database utilizing the DISTMIN calculation.

Speaker Recognition

Hidden Markov Processes are the measurable models in which one tries to describe the factual properties of the flag with the fundamental supposition that a flag can be portrayed as an arbitrary parametric flag of which the parameters can be evaluated in a précised and very much characterized way. In order to develop an isolated word recognition system using HMM, the following measures must be taken. A. For each expressed word, a Markov show must be assembled utilizing parameters that upgrade the perceptions of the word. B. Most extreme probability model is ascertained for the expressed word.

Limitations of Related works

1. The above told methods were failed to discuss some of the general features, and characteristic behavior of speech signals were not described properly.
2. The above discussed some of the tasks of speech recognition can be easily challenged by highly variant input speech signals.
3. In the presence of some noise, the above discuss process of handling devices can be overruled.
4. The above discussed methods, were failed to discuss about the use of filters for the noise removal.

WORKFLOW

1. Voice Recording MATLAB Code
2. Finally Voice recognition Code
3. GUI interfacing (if possible)

METHADOLOGY

Speech recognition is widely and mostly used in almost every security project where machine can recognize the person‘s voice as password to unlock it. For example, in user‘s daily life, if user want to turn the Geyser on or off using the voice commands then Speech Recognition plays a vital role. Application should understand the system and recognize the user commands ON or OFF.

Depending on limitations of other model, the technique called cross correlation for recognition of speech is used and simulated in MATLAB. Correlation compares the two signals, considering the five samples and comparing them with the test sample gives us the result. Every sound sample (test or five samples) are in .wav format. To recognize the words from the sound the concept Mel frequency cepstral coefficients (MFCCs) is used.

Speech Recognition

The nature of a speech recognition systems are evaluated by two elements: its accuracy (mistake rate in changing over talked words to advanced information) and speed (how well the product can stay aware of a human speaker). Speech recognition technology has unlimited applications. Generally, such programming is utilized for programmed interpretations, correspondence, sans hands figuring, restorative translation, mechanical autonomy, mechanized client administration, and a great deal more. On the off chance that you have ever paid a bill via telephone using an automated system, you have likely profit by speech recognition software.

Correlation Technique

Cross-correlation is a measure of similarity of two series as a function of the displacement of one relative to the other. This is also known as a sliding dot product or sliding inner-product. It is commonly used for searching a long signal for a shorter, known feature. It has applications in pattern recognition, single particle analysis. The term cross-correlation is utilized for alluding to the relationships between the sections of two arbitrary vectors X and Y, while the connections of an irregular vector X are thought to be simply the connections between simply the passages of X, those shaping the connection lattice (network of connections) of X.

xcorr function of MATLAB is an Cross-correlation function for sequence for a random process which includes autocorrelation.

Syntax for Correlation in MATLAB is derived as r = xcorr(x,y)

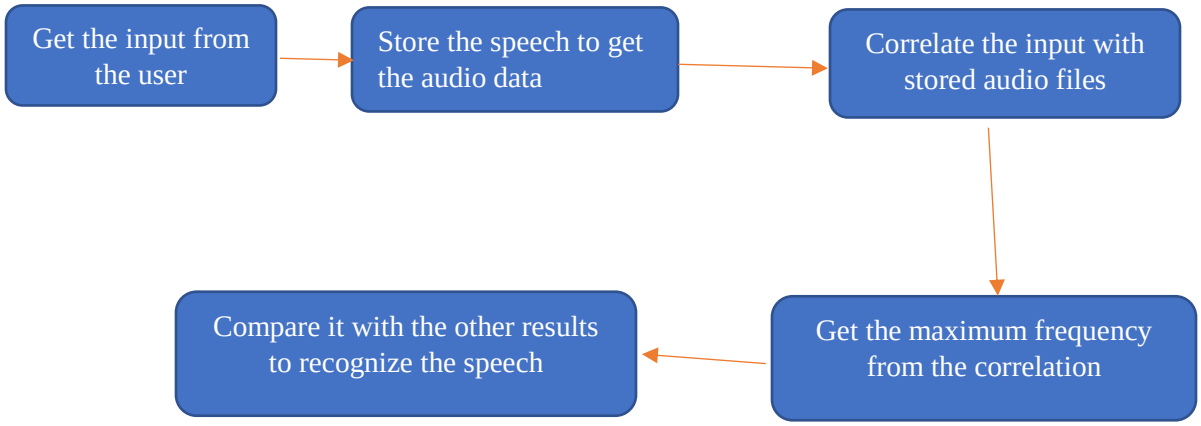
r = xcorr(x,y) returns the cross-correlation of two discrete-time sequences, x and y. Cross-correlation measures the closeness amongst x and moved (slacked) duplicates of y as a component of the slack. In the event that x and y have diverse lengths, the capacity annexes zeros toward the finish of the shorter vector so it has a similar length, N, as the other.

Correlation Technique

It is the representation of the short-term power spectrum of a sound, based on the linear transform of a log power spectrum. They are derived from a type of cepstral representation of the audio clip, this concept use the .wav format in MATLAB. MFCCs are commonly used as benefit in speech recognition systems which can automatically recognize the spoken words from the audio file. MFCCs are found in use of audio information retrieval applications such as genre classification, audio similarity measures. Its values are not very robust in the presence of additive noise, so it is easy to normalize their values in speech recognition systems to reduce the influence of noise.

CODE DESCRIPTION

The code which we used for our project is based on the work flow as described below:



Functions used in the code:

Audiodevinfo – To get the audio recording device information connected to the laptop of PC.

Audiorecorder - Audiorecorder creates an 8000 Hz, 8-bit, 1 channel audiorecorder object. A handle to the object is returned.

Recordblocking – To specify the time interval in which the audio input is sampled.

Getaudiodata – To get the data from the audio wave file.

Waveread – Retuns the sample rate FS in hertz and number of bits per sample

Xcorr – To get an estimated of the correlation between two random (jointly stationary) sequences.

Soundsc – To autoscale and play the vector as sound.

CODE

```
close all; clear all;

% get device information
dev = audiodevinfo;

% create recorder object
rec = audiorecorder(50000, 16, 1);

% start recording
disp('start speaking');
% record(rec, 5); % will record for 5sec. for this command pause needs to be added
recordblocking(rec, 5);

% stop recording
disp('Stop recording');

% Play recorded sound
play(rec);

y = getaudiodata(rec);
% plot the sound
figure
plot(y);

speechrecognition(y)
```

1. This is first part of our code. This code gets the audio device info of the system, then creates an audio recorder object with a specific frequency (“In our case it is 50000Hz”). The code asks the user to give an audio/voice input and records the speech for the next five seconds.
2. The recorded audio is then played and plotted on a graph and it calls the ‘speech recognition’ function, with given audio as the input for the next code.

Speech Recognition Code:

```
function speechrecognition(audio)

x=audio;
x=x';
x=x(1,:);
x=x';
y1=wavread('one.wav');
y1=y1';
y1=y1(1,:);
y1=y1';
z1=xcorr(x,y1);
m1=max(z1);
l1=length(z1);
t1=-((l1-1)/2):1:((l1-1)/2);
t1=t1';
subplot(3,2,1);
plot(t1,z1);

y2=wavread('two.wav');
y2=y2';
y2=y2(1,:);
y2=y2';
z2=xcorr(x,y2);
m2=max(z2);
l2=length(z2);
t2=-((l2-1)/2):1:((l2-1)/2);
t2=t2';
subplot(3,2,2);
plot(t2,z2);

y3=wavread('three.wav');
y3=y3';
y3=y3(1,:);
y3=y3';
z3=xcorr(x,y3);
m3=max(z3);
l3=length(z3);
t3=-((l3-1)/2):1:((l3-1)/2);
t3=t3';
subplot(3,2,3);
plot(t3,z3);

y4=wavread('four.wav');
```

```
y4=y4';
y4=y4(1,:);
y4=y4';
z4=xcorr(x,y4);
m4=max(z4);
l4=length(z4);
t4=-((l4-1)/2):1:((l4-1)/2);
t4=t4';
subplot(3,2,4);
plot(t4,z4);

y5=wavread('five.wav');
y5=y5';
y5=y5(1,:);
y5=y5';
z5=xcorr(x,y5);
m5=max(z5);
l5=length(z5);
t5=-((l5-1)/2):1:((l5-1)/2);
t5=t5';
subplot(3,2,5);
plot(t5,z5);

a=[m1 m2 m3 m4 m5];
m=max(a);
figure
plot(x);
h=wavread('allow.wav');
if m<=m1
    soundsc(wavread('one.wav'),50000)
        soundsc(h,50000)
        v = msgbox('You said one');
elseif m<=m2
    soundsc(wavread('two.wav'),50000)
        soundsc(h,50000)
        v = msgbox('You said two');
elseif m<=m3
    soundsc(wavread('three.wav'),50000)
        soundsc(h,50000)
        v = msgbox('You said three');
elseif m<=m4
    soundsc(wavread('four.wav'),50000)
        soundsc(h,50000)
        v = msgbox('You said four');
elseif m<m5
    soundsc(wavread('five.wav'),50000)
        soundsc(h,50000)
        v = msgbox('You said five');

else
   v = msgbox('Speech not recognised');

end
```

This code is a function that we have created in order to perform speech recognition.

The first part of the code gets the audio input provided from the user by the record audio code and store it in a variable.

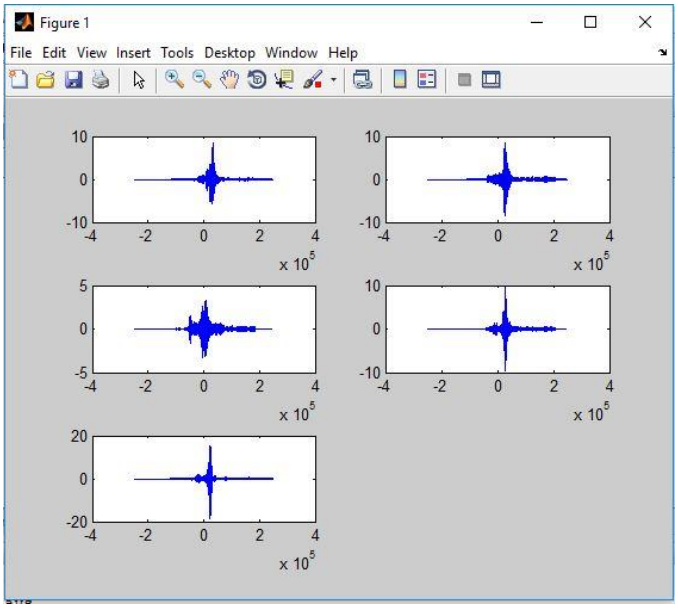
Then, it performs the correlation of the given input with other audio files that are already stored in the database and plots the results.

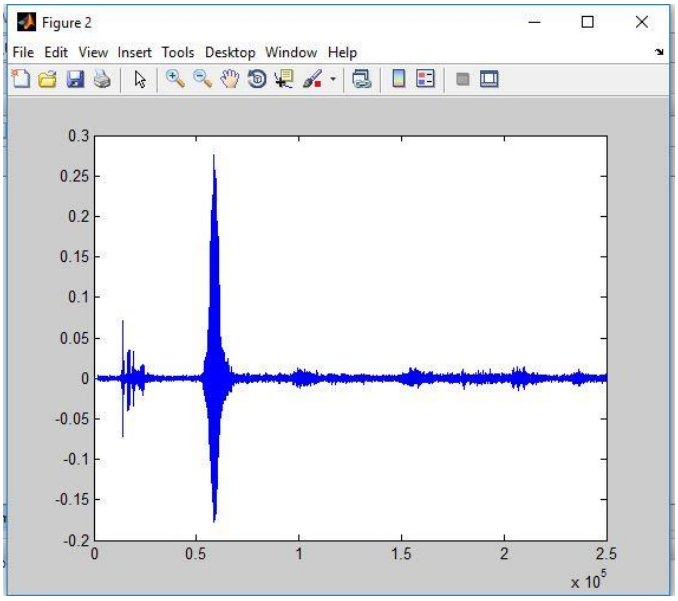
Next, the code takes the maximum frequency from each result and stores them in an array.

Then the maximum value in that array is again compared with all the elements of the array to tell which is maximum.

Finally, the corresponding audio file is selected as the result and the audio is played as the output.

OUTPUT





OTHER APPLICATIONS

1. Solving Crimes with Voice Recognition: When it comes to a crime scene, there are some forensics situations in which only audio evidence is available to investigators, and that's where voice biometrics can be deployed to great effect. Such speech recognition can provide great help in nabbing criminals.

CONCLUSION

This project defines us successfully about various features, behavior and characteristics of speech signals and also deals with the concept of cross correlation. In this project, an algorithm has been created with the help of MATLAB programming which requires .wav format speech input signals where comparison with the test sound file using correlation technique takes place. Thus, project concludes that in order to remove the further limitation of audio formats there is a requirement for the study of various formats of speech signals which will be further used for communication with the machines which include the hardware part and not the simulator.

REFERENCES

[1] Pramanik, A., & Raha, R. (2012, October). Automatic Speech Recognition using correlation analysis. In *Information and Communication Technologies (WICT), 2012 World Congress on* (pp. 670-674). IEEE.

[2] Gupta, A., Raibagkar, P., & Palsokar, A. Speech Recognition Using Correlation Technique.

[3] Shaneh, M., & Taheri, A. (2009). Voice command recognition system based on MFCC and VQ algorithms. *World Academy of Science, Engineering and Technology*, *57*, 534-538.

[4] 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.

THANK YOU