WALCHAND COLLEGE OF ENGINEERING, SANGLI

(Government-Aided Autonomous Institute)

Department of Electrical Engineering

'Digital Signal Processing'

(Mini Project Synopsis)

"Voice Biometric using Speech Recognition "

Submitted By :EL-4 group 4

1. **Mr. Sandesh Martand Mane :2019BTEEL00078**
2. **Mr. Prathamesh Anandrao Shinde :2019BTEEN00014**
3. **Mr. Rohit Bhat : 2020BTEEL00201**
4. **Miss. Sapana Dilip Uike : 2020BTEEL00202**
5. **Mr. Pranav Laxman Sadafal : 2020BTEEL00203**

Under the guidance of

"A.B. Patil Sir

TABLE OF CONTENTS

ABSTARCT

Speech is a prominent communication method among humans, with the development Of multimedia technology, speech recognition technology has increasingly become a hotspot of research in recent years. It has a wide range of applications, which deals with recognizing the identity of the speakers that can be classified into speech identificationand speech verification according to decision modes. Speech recognition is used in almost every security project where you need to speak and tell your password to computer and is also used for automation.

The main work of this project is to study and research the techniques, algorithms of speech recognition, thus to create a system to simulate the speech recognition This Project also demonstrates a model that shows technological advancement where humans and computers interact via voice user interface. In developing the model, cross correlation was implemented in MATLAB to compare two or more signals and detect the most accurate one of the all. We are actually used cross correlation to find similarity between our recorded Signal files and the testing signal. Thus we were able to develop a model where machines can differentiate between commands and act based on it.

INTRODUCTION

Voice is already the gateway to many services and aplications that have become an integralpart of our daily lives. The level of user-friendliness enabled by voice-based interfaces place them heads and shoulders above other options when it comes to interacting in a seamless way with our devices. From a biometrics standpoint, voice offers additional benefits.

It is a lot easier for us humans to recognize a person's voice than computers. Hence, speech recognition in machine learning is a game changer as developing machines that can understand and uniquely identify a person's voice would make Human-Computer interaction more intriguing. Speech recognition is one of the next generation technologies for human-computer interaction.

Speech recognition has been researched since the late 1950s but due to its computational complexity and limited computing capabilities of the last few decades, its progress has been hindered. In laboratory settings automatic speech recognition systems

(ASR) have achieved high levels of recognition accuracies, which tend to degrade in real world environment. In today era speech technologies play an important role. This technology is commercially and easily available for a different uses. These technologies make machines
respond correctly and it provides valuable services. In modern era, no one wants to reveal his identity due to security purposes because every person has different speech characteristic. Thus with the different information in speech waves we can easily identify the speaker.

WORK OBJECTIVES :

The Main Aim of the Project is to Recognize the sample Voice provided to the System by use of cross correlation Method .

LITERATURE REVIEW

Speech recognition methods can be divided into textindependent and text dependent methods. In a text independent system, speaker models capture characteristics of somebody's speech, which show up irrespective of what one is saying. In a text-dependent system, on the other hand, the recognition of the speaker's identity is based on his or her speaking one or more specific phrases, like passwords, card numbers, PIN codes, etc. [3].Every speech recognition application is designed to accomplish a specific task. Examples include: to recognize the digits zero through nine and the words "yes" and "no" over the telephone, to enable bedridden patients to control the positioning of their beds, or to implement a VAT (voiceactivated typewriter). Once a task is defined, a speech recognizer is chosen or designed for the task [2]The task of speech recognition is to convert speech into a sequence of words by a computer program.

Speech recognition applications enable people to use speech as another input mode to interact with applications with ease and effectively. Speech recognition interfaces in native language will enable the illiterate/semi-literate people to use the technology to greater extent without the knowledge of operating with computer keyboard or stylus [4].

There are different modes available for Speech Recognition System:

Speaker Dependent / Independent System: It must be trained in order to recognize accurately what has been said. To train a system, Speaker is asked to record predefined words or sentences that will be analyzed and that results will be stored.

Word Recognition: It is Simplest mode and less greedy in terms of CPU requirement. Word is surrounded by silence so that boundaries are well known.

Continuous Speech Recognition: It assumes that system is able to recognize a sequence of words in a sentence.

Keyword Spotting: It is able to identify in a sentence a word corresponding to a particular command. Created to cover the gap between isolated and continuous System. Vocabulary Size: Larger the vocabulary the system can make more errors. So vocabulary size matters.

METHODOLOGY

Five recorded wav sound samples were stored in the database and we wish to recognize them using a correlation test.wav in MATLAB.
STEP1: We create a test file command as shown in figure 1.

```
clc
clear all
close all
warning off
Fs=8000;
ch=1;
datatype='unit8';
nbits=16;
Nseconds=3;
recorder=audiorecorder(Fs,nbits,ch);
disp('Start Speaking');
recordblocking(recorder,Nseconds);
disp('End Speaking');
x=getaudiodata(recorder,datatype);
audiowrite('one.wav',x,Fs);
```

Fig.1

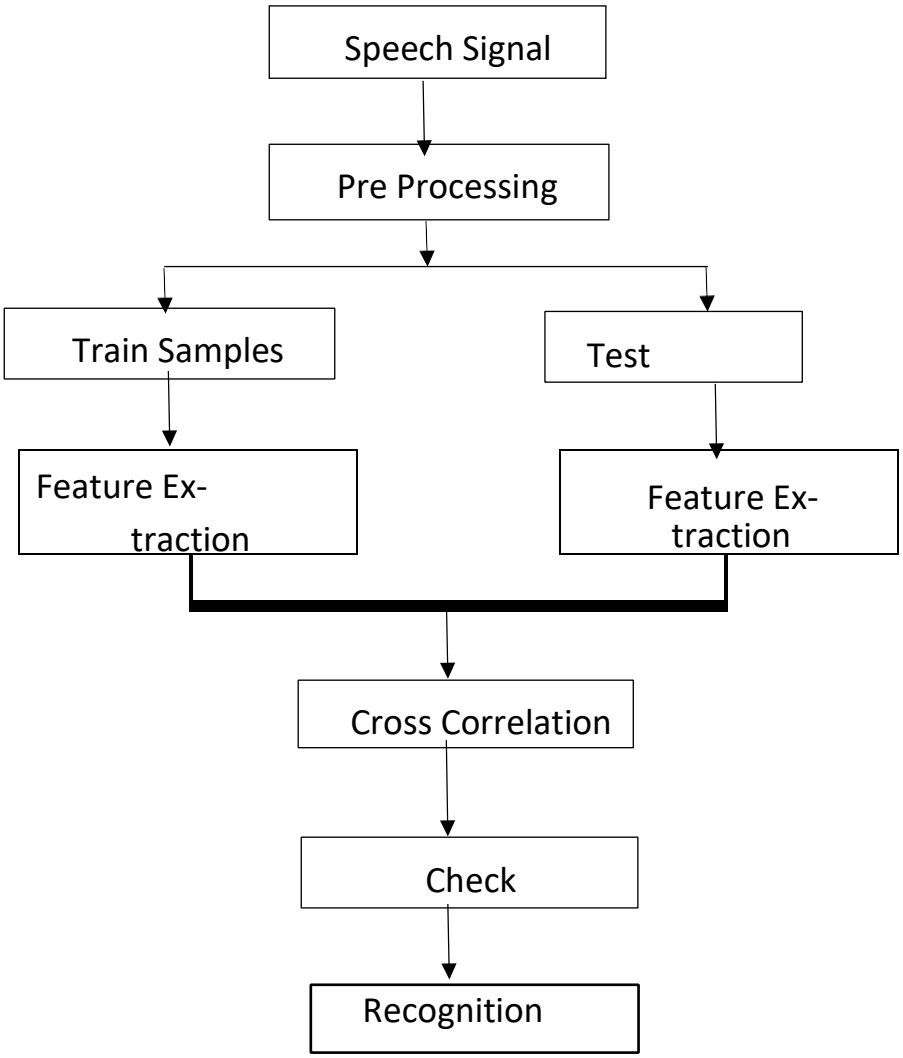
STEP2: The code shown in figure 2 represents the sample one.wav, do this repeatedly for as many samples you want to be present. In this case, we have five samples. Therefore, we repeated this code five times with every place that say 1 replaced with 2, 3, 4 and 5 respectively.

STEP3: We note that the test has x value, it is this value that is used to compare the y value of the sample. Hence, Z1=xcorr(x,y);

STEP4: We create a conditional statement where m6=300; and if m=max the machine will sound the allowed

```
plot(t1,z1);
y2=audioread('2.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);
figure
plot(t2,z2);
y3=audioread('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)
figure
plot(t3,z3);
y4=audioread('four.wav');
y4=y4';
y4=y4(1,:);
y4=y4';
z4=xcorr(x,y4);
m4=max(z4);
l4=length(z4);
```

FLOW CHART


Speech Signal
Pre Processing
Train Samples
Test
Feature Ex-
traction
Feature Ex-
traction
Cross Correlation
Check
Recognition

PROJECT OVERVIEW

In this project we are going to demonstrate the Biometric Voice Recognition in MATLAB using Cross-Correlation method, where we are going to first generate and save 5 sample voices of 5 participants using MATLAB code to create train samples of voice and we are going to save it in '.wav' format.

Secondly by using the same code we are going to create one test sample of any random participant. Further by we are going to implement the code of cross-correlation, where we compare the test sample with all the train sample to get the graphs of each correlation. Then by examination of graphs we can find that out of which train sample matches to test sample.

CODE :

```
clc
clear all
close all
warning off
Fs=8000;
ch=1;
datatype='unit8';
nbits=16;
Nseconds=3;
recorder=audiorecorder(Fs,nbits,ch);
disp('Start Speaking');
recordblocking(recorder,Nseconds);
disp('End Speaking');
x=getaudiodata(recorder,datatype);
audiowrite('one.wav',x,Fs);
```

- A UINT8 is an 8-bit unsigned integer (range: 0 through 255 decimal). Because a UINT8 is unsigned, its first bit (Most Significant Bit (MSB)) is not reserved for signing.

- recordblocking(recorderObj , length) records audio from an input device, such as a microphone connected to your system, for the number of seconds specified by length . The recordblocking method does not return control until recording completes.

- y = getaudiodata(recorder) returns recorded audio data associated with audiorecorder object recorder in a double array y . y = getaudiodata(recorder , dataType) converts the signal data to the specified data type.

- Description. audiowrite(filename , y , Fs) writes a matrix of audio data, y , with sample rate Fs to a file called filename . The filename input alspecifies the output file format. The output data type depends on the output file format and the data type of the audio data, y.

```
function speech recognization(filename)

%speech recognization using correlation method
 %speechrecognization('sample.wav')
 voice=audioread(filename);
 x=voice;
 x=x';
 x=x(1,:);
 y1=audioread('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=audioread('2.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);
 figure
 plot(t2,z2);
 y3=audioread('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)
 figure
 plot(t3,z3);
```

RESULT OF TESTED VOICE SAMPLES:

In the command window, we typed the command “speechrecognition (‘test.wav’) and hit the enter button to get the spectrum graphs. In this particular case, test.wav and test2.wav represents the number 2 and 3 respectively. So we run a test for test2.wav. Now because test2.wav sounds the number 3, figure 3 as shown in figure 4 is the most accurate spectrum. We repeated the step for test.wav which represents number 2. From figure 5 we can see that graph 2 is more accurate as test.wav sound is two.

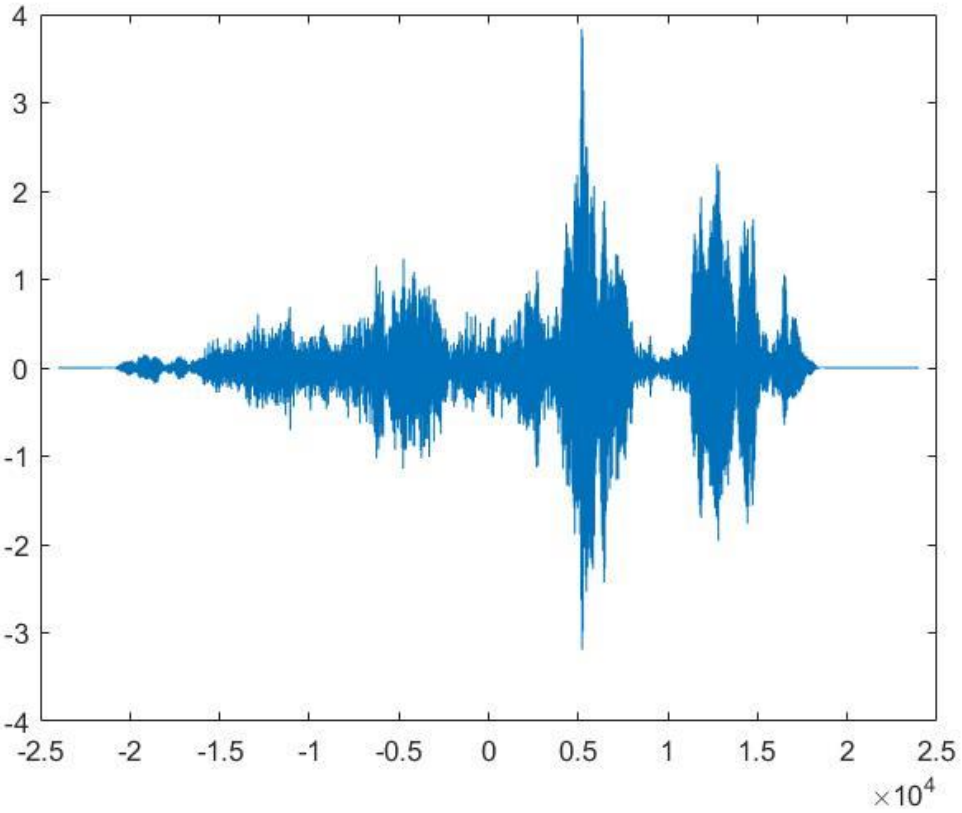


Fig.1

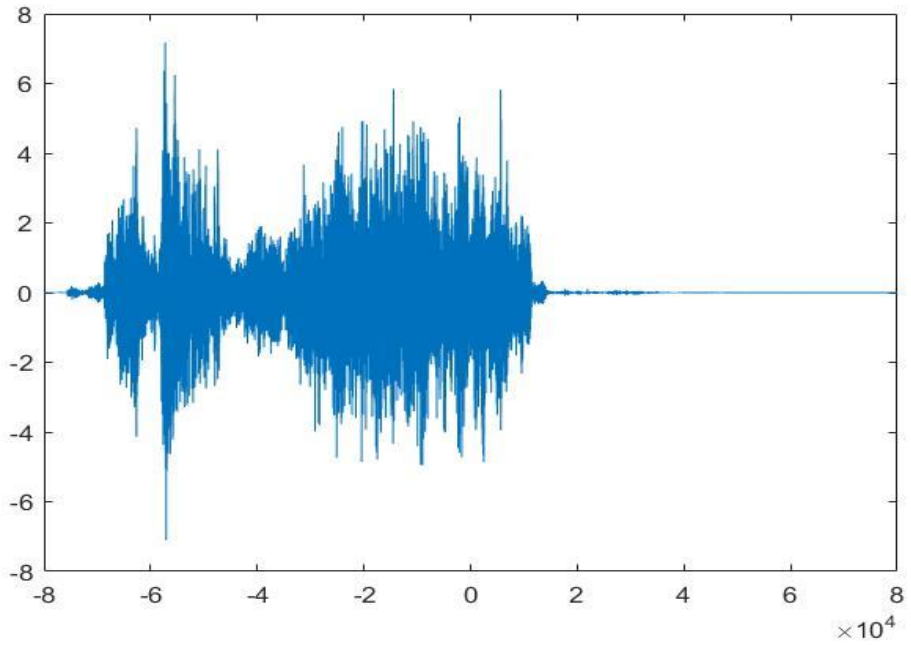

8
6
4
2
0
-2
-4
-6
-8
-8
-6
-4
-2
0
2
4
6
8
×10^4

Fig.2

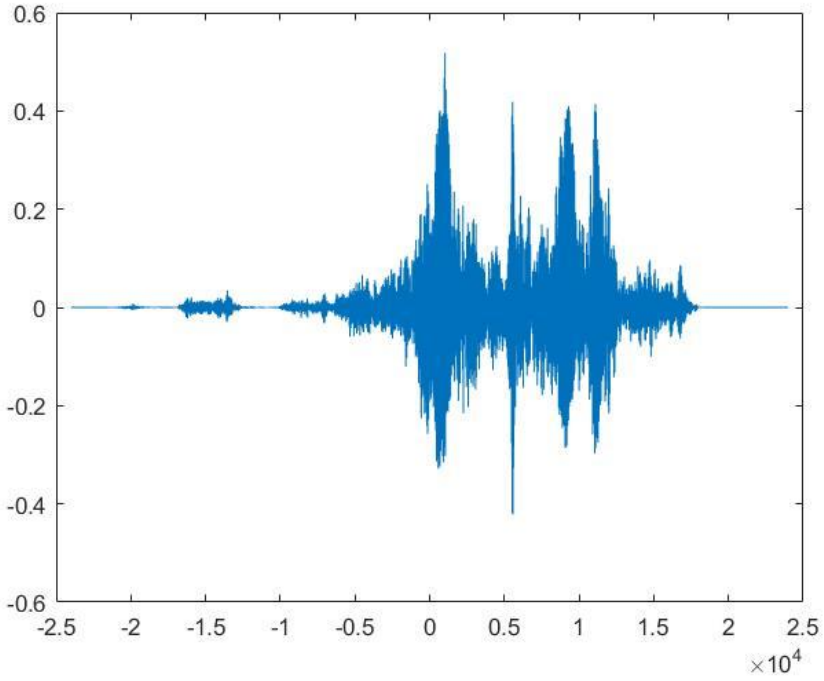

0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
×10^4

Fig.3

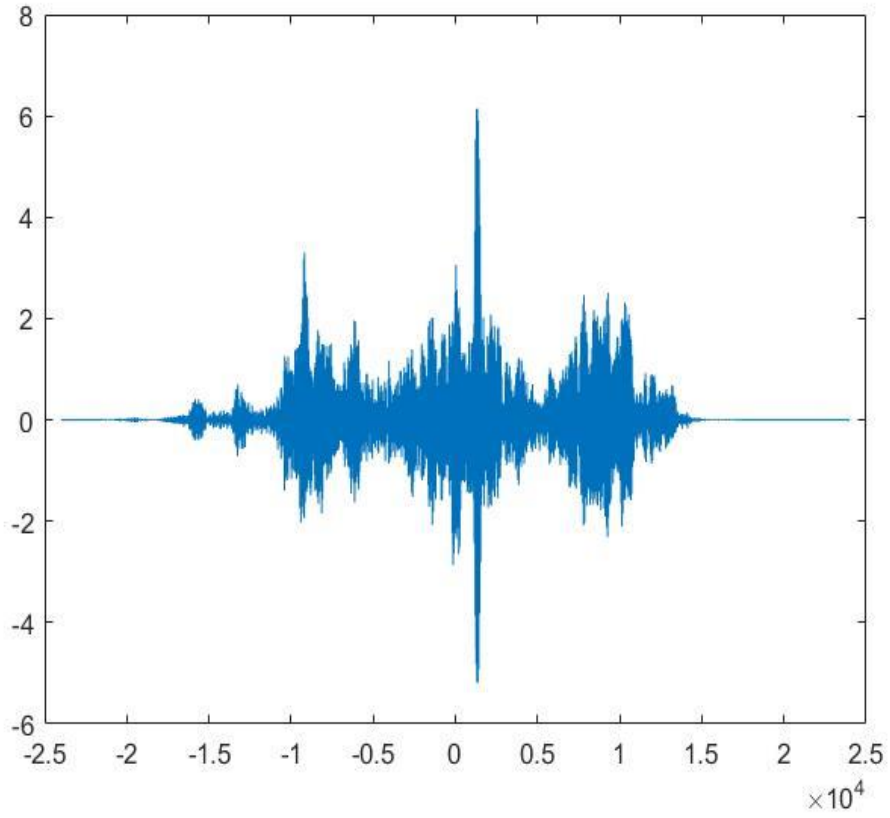

8
6
4
2
0
-2
-4
-6
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
×10^4

Fig.4

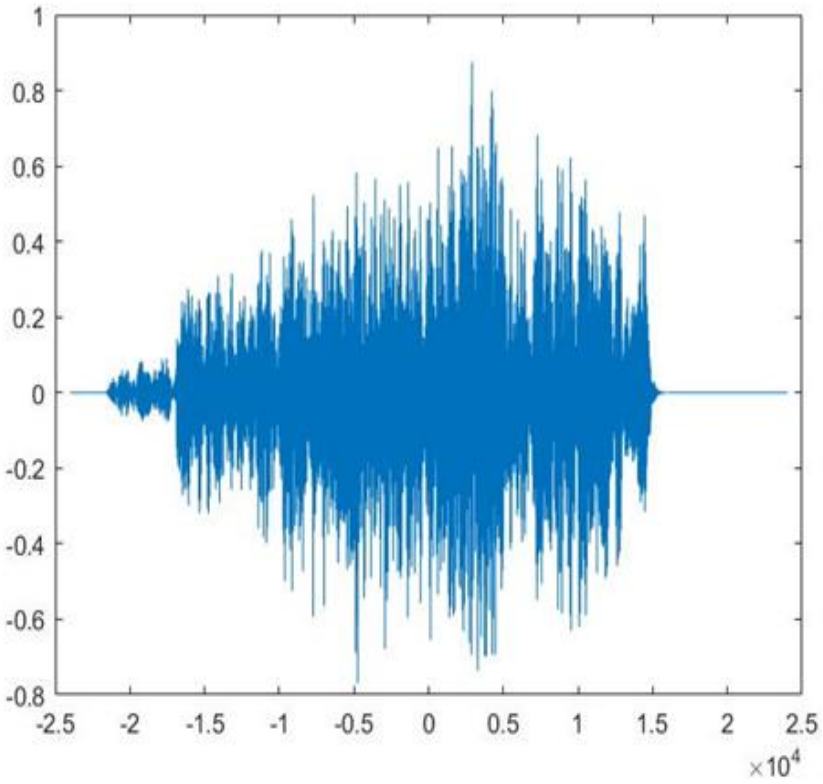

1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
×10^4

Fig.5

ADVANTAGES:

1. Voice commands are a far more efficient tool than typing a message.
2. Voice recognition and speech activation is being developed for a whole myriad of reasons.
3. Lower operational cost.
4. Voice can communicate emotions.
5. It reduced cost and error.
6. This technology solves inefficiencies and reduces wasted time.
7. Speech recognition may be possible to reduce overtime hours for transcriptionists and/or stop outsourcing clinician dictation to outside firms.
8. It is more efficient for everyone to save money in the process.

DISADVANTAGES:

1. Difficult to build a perfect system.
2. Speaking is loud and invites noise to others.
3. Filtering background noise is a task and it is too much that can even be difficult for humans to accomplish.
4. This biometric is sensitive to environmental conditions such as background noise.
5. It doesn’t keep privacy it does not suit a crowded environment.
6. Error and misinterpretation of words.

APPLICATION:

1)Voice biometrics for security: Voice biometrics use a person's voice as a unique identifying biological characteristic in order to authenticate them. Speech recognition can also be used for voice authentication to replace processes where a user has to display her personal information to authenticate herself.

2)Healthcare: (Diagnosis):

Machines can estimate a person's mental state by analyzing his/her voice. For example, such
models can estimate whether a patient is depressed or suicidal.

3)Industry applications: (Automotive):

In car speech recognition systems have become a standard feature for most new vehicles. These systems aim to remove the distraction of looking down at your mobile phone while you drive. Thanks to these systems, drivers can use simple voice commands to initiate phone calls, select radio stations or play music.

4)Voice commands to smart home devices:

Smart home applications are mostly designed to take a certain action after the user gives voicecommands. Smart home devices are widely used speech recognition application.

CONCLUSION :

The voice recognition algorithm is developed by cross correlation method to extract the feature of the voice signal. The reference voice is being stored in training phase and compare with voice in testing phase to match both the results. The system is successfully recognize rhe authenticate user's voice and reject all other imported voices.

FUTURE SCOPE :

A few recommendations are provided to improve the accuracy and the performance of the voice recognition system. Firstly, increase the accuracy of voice recognition system. The background noise must be filtere completely to get more accurate data.Increase the complexity of the voice recognition by limit the range of the amplitude or frequency of the voice signal. So the system can be recognize the admin's voice more accurate.