Institute Of Technology
NIRMA UNIVERSITY

PROJECT REPORT
ON

SPEECH RECOGNITION USING CORRELATION IN MATLAB

Submitted by

DARSHIL MAVADIYA (22BEC508)
SAKSHI LIKHIA (20BEC060)

Guided by
DR.RUTUL PATEL
DR.BHUPENDRA FATANIYA

DIGITAL SIGNAL PROCESSING (2EC502)

In fulfillment for the Course of
BACHELOR OF TECHNOLOGY
in
Electronics and Communication
2023-24

ABSTRACT

Speech is a prominent communication method among humans, whereas the communication between human and computers were based on text user interface and graphic user interface. 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. This paper demonstrates a model that enhances 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 upon them.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

- Speech Recognition is the way of capturing the talked words using a gadget and converting them into a digitally stored set of words. 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. In the current world, there is a continually expanding need to confirm and recognize the voice of individuals automatically.
- Speech recognition is basically and widely used concept for providing the security to the applications. Security has become a major part for any user using any smart devices. Speech Recognition is one of part of Biometrics. Biometrics, the physical qualities and behavioral attributes that make each of us exceptional, are a characteristic decision for personality confirmation.
- 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. So, Speech can be used for the identification of person because every person has different speech characteristic. Thus with the different information in speech waves we can easily identify the speaker.
- Speech Recognition is the way of capturing the spoken words using a gadget and converting them into a digitally stored set of words. In the current world, there is a continually expanding need to confirm and recognize the voice of individuals automatically. Speech recognition is a basic and widely used concept for providing security to the applications.

.

CHAPTER 2: LITERATURE SURVEY

2.1 What Is The Speech Recognition ?

- Speech recognition is an interdisciplinary subfield of computer science and computational linguistics that develops methodologies and technologies that enable the recognition and translation of spoken language into text by computers with the main benefit of searchability. It is also known as automatic speech recognition (ASR), computer speech recognition or speech to text (STT). It incorporates knowledge and research in the computer science, linguistics and computer engineering fields. The reverse process is speech synthesis.
- The term voice recognition or speaker identification refers to identifying the speaker rather than what they are saying. Recognizing the speaker can simplify the task translating speech in systems that have been trained on a specific person's voice or it can be used to authenticate or verify the identity of a speaker as part of a security process.
- From the technology perspective, speech recognition has a long history with several waves of major innovations. Most recently, the field has benefited from advances in deep learning and big data. The advances are evidenced not only by the surge of academic papers published in the field, but more importantly by the worldwide industry adoption of a variety of deep learning methods in designing and deploying speech recognition systems.

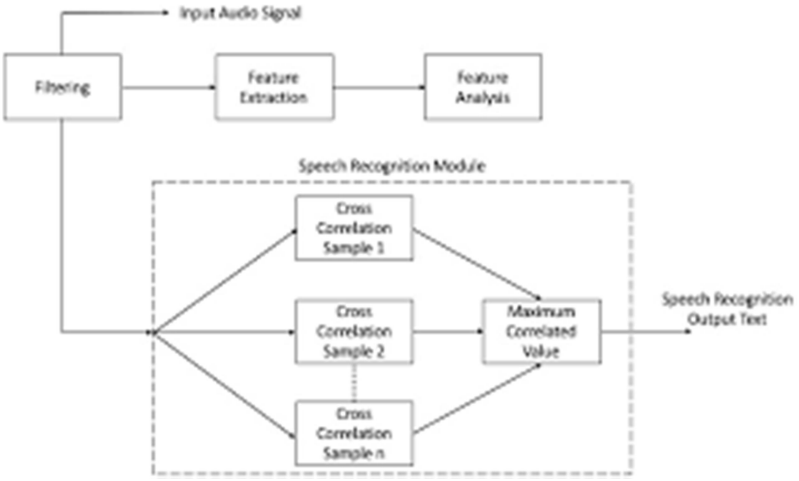


Fig. 1 Block Diagram of speechrecognition using correlation

2.2 Correlation:

Correlation is a measure of similarity between two signals, indicates the measure up to which One signal resembles the other signal. Correlation is a mathematical operation which is similar to convolution. Just as with convolution, correlation uses two signals to produce a third signal. Co-variance is the measure of the deviation between two sets of random variables. It compares the two signals, By considering the provided samples and comparing it with the test sample to provide the result. Syntax for Co-variance in MATLAB is derived as: **r = cov(x)**

r = cov(x) returns a specific covariance value to the program which can be compared with the other signals for speech recognition. Covariance method measures the difference between the covariance value of the sample voice to that of the value of predefined voice and then gives the result whether accepted or not.

If the voice is recognized, it returns an electric sound, otherwise returns "Access Denied" sound if the given sample does not match.

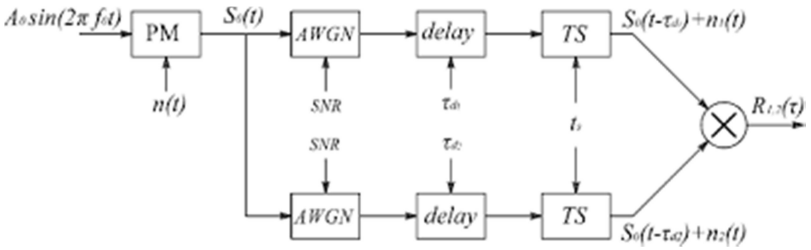


Fig. 2 Block diagram of correlation

CHAPTER 3: PROBLEM DESCRIPTION

3.1 Key points of problem description in speech recognition :

Hence, what is the need of speech recognition?

Speech recognition is essential for improving accessibility, increasing efficiency, enhancing user experiences, and enabling automation in various fields and applications. It is a versatile technology with a wide range of uses, making it a valuable tool in our increasingly technology-driven world. This technology is a process of extracting the speech characteristic information from people's voice, and then being operated through the computer and to recognize the content of the speech. But the voice characteristic parameters of different people are different, such as the loudness, voice amplitude, pitch etc. The term speech recognition refers to identifying the speaker, rather than what they are saying.

So, we build this program in matlab using correlation technique. The input signal is correlated with each of the sample files present in the matlab database, and if a correspondence is found, it grants access or lese it will deny the access.

CHAPTER 4: METHODOLOGY

4.1 Methodology of Speech Recognition :

Correlation uses two signals to produce a third signal. It is the measure of similarity between two signals. By analyzing the correlation between the input and reference samples, the system can determine if the speaker is known or unknown.

Then we use co-variance. It is the measure of the deviation between two set of variables. Here, it compares the two signals. Syntax for co-variance in MATLAB is derived as r = cov(x), it returns a specific covariance value to the program which can be compared with the other signals for speech recognition.

Lastly, if the voice is recognized, it returns an electric sound, otherwise returns "Access Denied" sound if the given sample does not match

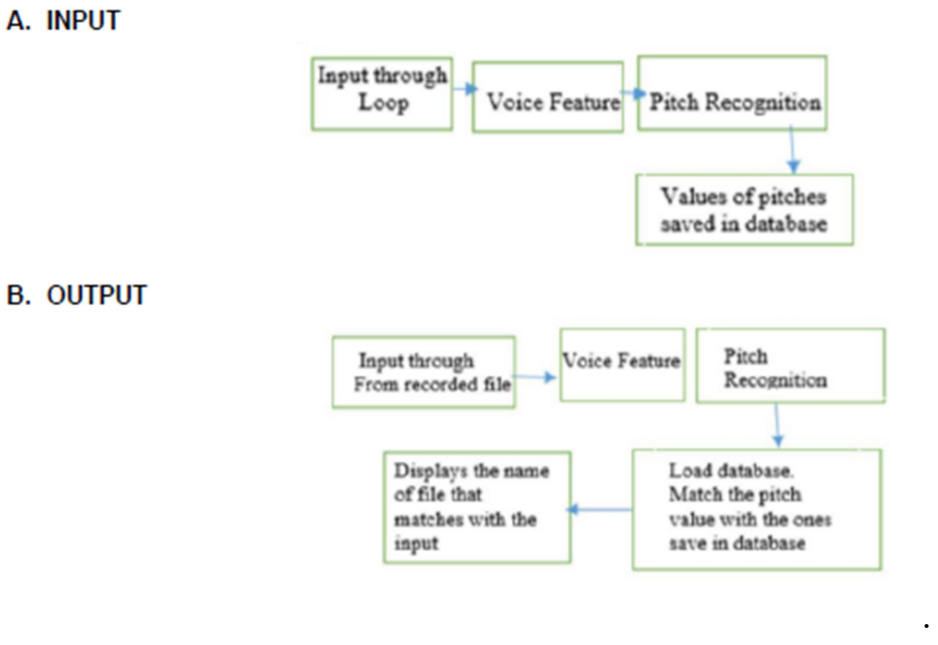
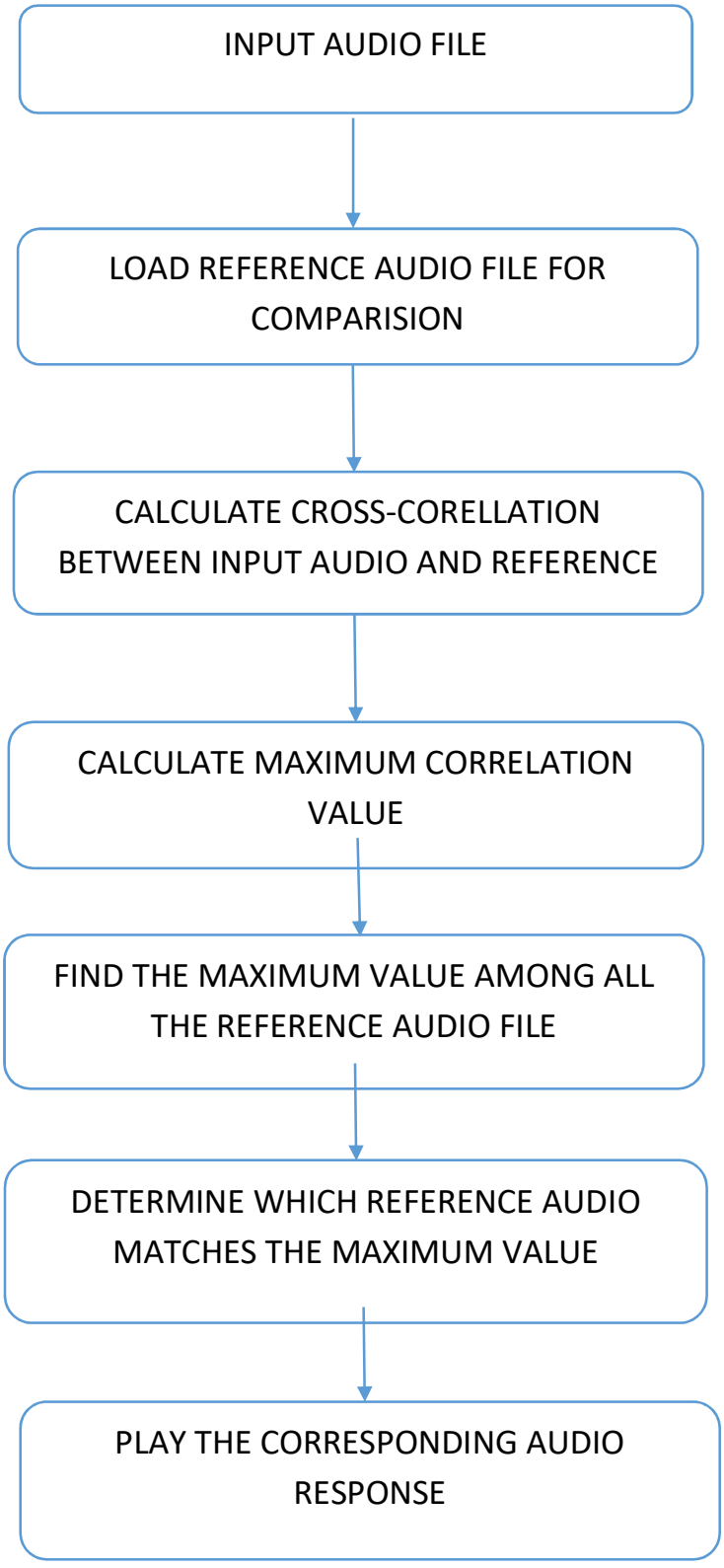


Fig.3 Block diagram of methodology

CHAPTER 5: FLOWCHART

5.1 Flowchart of speechrecognition:



CHAPTER 6: ALGORITHM

6.1 Algorithm of speechrecognition:

1. It loads an input audio file (specified as 'test.wav') and reads its data.
2. The script then loads reference audio files ('one.wav', 'two.wav', 'three.wav', 'four.wav', 'five.wav') for comparison. Each reference audio file corresponds to a specific spoken word or number ('one', 'two', 'three', 'four', 'five').
3. For each reference audio file, the code computes the cross-correlation between the input audio and the reference audio using xcorr() and stores the result in variables like z1, z2, z3, etc. It also determines the maximum correlation value (m1, m2, m3, etc.) and the length of the correlation output.
4. The code then plots the correlation results using the plot() function for visual inspection.
5. It calculates the maximum correlation value among all reference words (m).
6. If the maximum correlation value (m) corresponds to one of the reference words ('one', 'two', 'three', 'four', 'five'), it plays the audio of that word using the soundsc() function. If the maximum correlation doesn't match any reference word, it plays 'denied.wav' as the output.
7. Use of soundsc function is used. Syntax: soundsc(y,Fs) For example, Play an excerpt from Handel's Hallelujah Chorus at twice the recorded sample rate

   load handel.mat

   soundsc(y,2*Fs)
8. As a result, figures will be generated and if the correlation of the input signal matches the correlation between any of the audio samples and the input signal, then they are the same meaning that the input audio and the samples are equivalents. I.e they are recognized

CHAPTER 7: PSEUDO CODE

7.1 Pseudo code of speechrecognition using correlation :

- In order to start running the code, the user has to insert the command written in line 2 in the command window. Before pressing enter, the user should choose which one of the test files we would like to try (the audio file's name are: one,two,three,four, with the first ones granting access, and the latter denying access by using if-else ladder).After doing what is above stated and then pressing enter, the code will start to run.On line 2, by using the command audioread, the code reads data from the file names between the parenthesis, and returns sampled data. This will give the program the necessary information to start an autocorrelation on line 11.
- After doing so, the program will do the similar process with the 5 audio samples that are designed to have access. Instead of an autocorrelation, it will perform a cross-correlation between each of the 5 audio samples and the input signal. Such a process will occur from line 7. Within this process, some other tasks are completed.The Program will extract the peak value of each of the cross-correlations for future use and also plot the waveforms of each of the resulting correlations done.
- After that, a vector shall be formed with each of peak value of audio samples(line 67) And the maximum value of peak value will be extracted(line 68).
- Following the above-mentioned process, the program will enter in its final stage, the comparison and decision stage. From line 70 to 88, the program will compare the Maximum peak value f the vector from line 67 with the peak value of each of the Cross-correlations.
- If there is a match with one of them, the program will grant access And reproduce a sound file conveying such information with use of command soundsc, If not it will deny access, and also reproduce a sound file conveying such information.

CHAPTER 8: MATLAB CODE

8.1 Speech Recognition Using Correlation MATLAB code:

```
function speechrecognition(~)
%Speech Recognition Using Correlation Method
voice= audioread('test.wav');
%recObj = audiorecorder;
%disp('Start speaking.');
%recordblocking(recObj, 4); % Record for 5 seconds
%disp('End of Recording.');
%inputAudio = getaudiodata(recObj);
%voice = inputAudio;
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('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);
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);
t4=-((l4-1)/2):1:((l4-1)/2);
t4=t4';
%subplot(3,2,4);
figure
plot(t4,z4);
y5=audioread('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);
figure
plot(t5,z5);
m6=300;
```

```
a=[m1 m2 m3 m4 m5 m6];
m=max(a);
h=audioread('allow.wav');
if m<=m1
soundsc(audioread('one.wav'),50000)
soundsc(h,50000)
elseif m<=m2
soundsc(audioread('two.wav'),50000)
soundsc(h,50000)
elseif m<=m3
soundsc(audioread('three.wav'),50000)
soundsc(h,50000)
elseif m<=m4
soundsc(audioread('four.wav'),50000)
soundsc(h,50000)
elseif m<m5
soundsc(audioread('five.wav'),50000)
soundsc(h,50000)
else
soundsc(audioread('denied.wav'),50000)

end
```

CHAPTER 9: RESULT/ OBSERVATIONS

9.1 All Over Result:

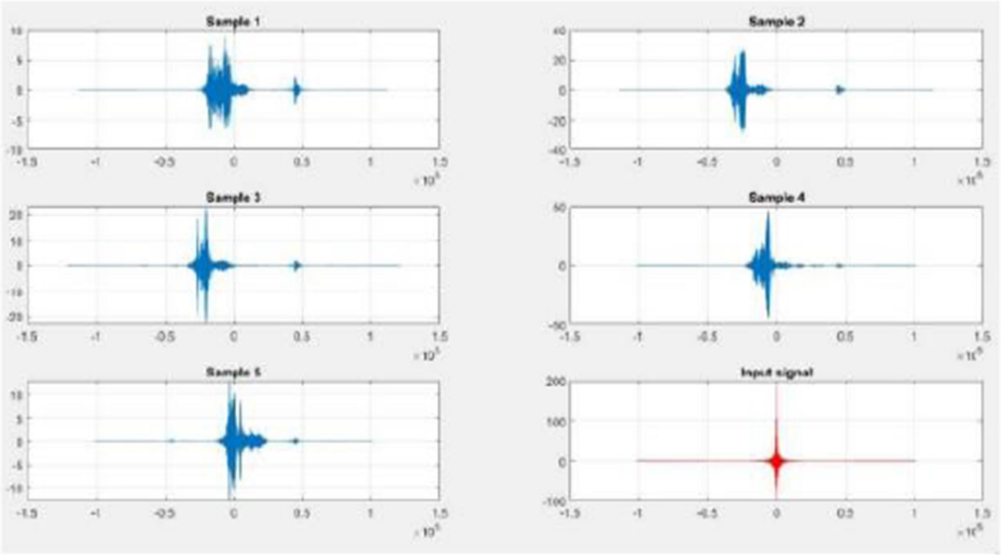


Fig. 4 Observation of all wave

- As it can be seen from the figure, the autocorrelation of the input signal does not match the cross-correlation between any of the audio samples and the input signal are the same,meaning that input audio is not equivalent to any of the audio samples.
- Using MATLAB, the graphs for comparison between the test and sample audio files is derived. Two test files and five sample files which has audio (Spoken word) of one to five is considered. One test file is match from five sample files and another test file is the denied file which is not matched with any sample files. When a test file is given as the input, the loop starts where first the spoken word from the audio files are computed and correlated with each other and using MALAB the graph where frequency of speech is displayed. Let‘s consider the test.wav file which is match for the second sample. When the input speechrecognition(‘test.wav’) is given in MATLAB, the comparison will start.
- Now consider the denied.wav file which is not a match with any sample given. When the given input speechrecognition(‘denied.wav’) in MATLAB command

prompt, the comparison will start and it will tell denied which means the file is not matched with any of the sample files.

The below are the graphs:

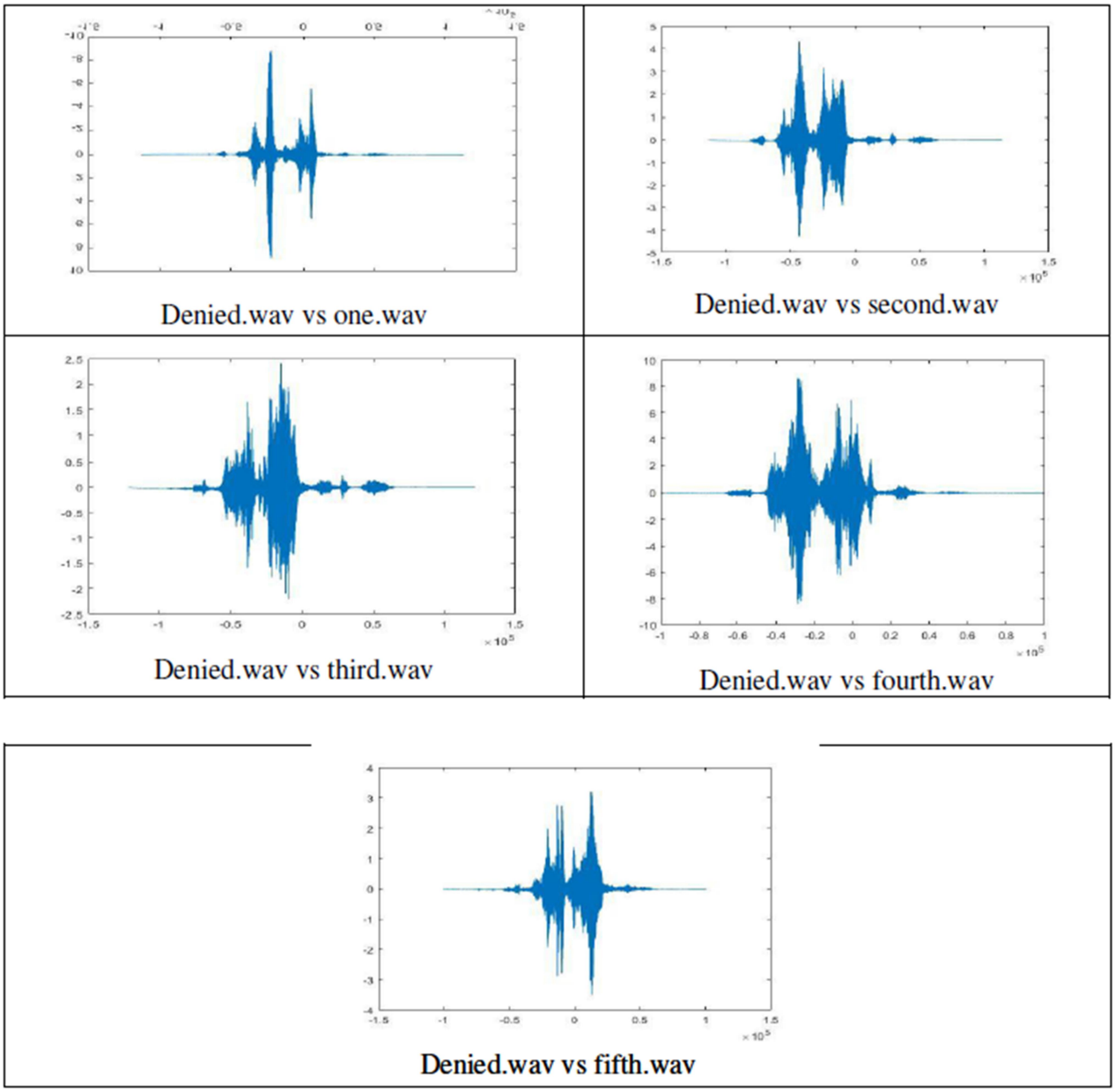


Fig. 5 Observation of Denied wave

- When we see in the success result, second sample is the successful match so at coordinates (0, 0) the words of audio file match which is seen in frequency format in graph.

CHAPTER 10: CONCLUSION

This code provides a basic illustration if speech recognition using a correlation method. It compares the input audio with reference audio files for known words. The recognized word is based on the highest correlation value. More advanced techniques are necessary for practical speech recognition systems.