A Report on

**Speech recognition using Correlation**

For

**Course Project(REV-2019-’C’ Scheme)of Second Year, (SE Sem-IV)**

in

**Department of Electronics & Telecommunication Engineering**

by

| **Roll No.** | **Name of Student** |
| --- | --- |
| 02 | Divya Mokal |
| 14 | Aditi Patil |
| 15 | Kanchan Patil |
| 20 | Tejas Pote |

Under the guidance of

**Prof. Sweta Komalkar**

****

**UNIVERSITY OF MUMBAI**

**A. Y. 2020-21**



**Mahatma Education Society’s**

**Pillai College of Engineering**

**Accredited A+ by NAAC**

**Dr. K. M. Vasudevan Pillai Campus,**

**Plot No. 10, Sector-16, New Panvel - 410206**

**AIM:** Toimplement cross correlation in MATLAB and to compare two or more signals and detect the most accurate one of them all.

**SOFTWARE USED:** MATLAB

**THEORY:**

**Introduction:**

Speech is a prominent communication method among humans, whereas the communication between humans and computers was based on text user interface and graphical user interface. Speech recognition is used in almost every security project where you need to speak and tell your password to a computer and is also used for automation.

Speech is the most prominent means of communication amongst humans. Human-to-human interaction is based on speech, emotion and gestures, thereby making it a lot easier to understand one another. On the other hand, the communication between humans and computers is based on either Text User Interface (TUI) or Graphic User Interface(GUI). 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.

In today's era speech technologies play an important role. This technology is commercially and easily available for a different use. These technologies make machines respond correctly and it provides valuable services. In the modern era, no one wants to reveal his identity due to security purposes. So, Speech can be used for the identification of a person.

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.

**Cross correlation:**

Cross correlation is a standard method of measuring the similarities/relationships between two signals. It is a measure of similarity of two series as a function of the displacement of one relative to the other. There are some cases where it is necessary to compare one reference signal with one or more signals to determine the similarities between signals and to determine additional information based on their relationships. 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, electron tomography, averaging, cryptanalysis, and neurophysiology. 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. In MATLAB the cross correlation function is xcorr for sequence for a random process which includes autocorrelation. Thus the 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.

**PROGRAM:**

function speechrecognition(filename)

%Speech Recognition Using Correlation Method

%Write Following Command On Command Window

%speechrecognition('test.wav')

voice = audioread(filename);

x=voice;

x=x';

x=x(1,:);

x=x';

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

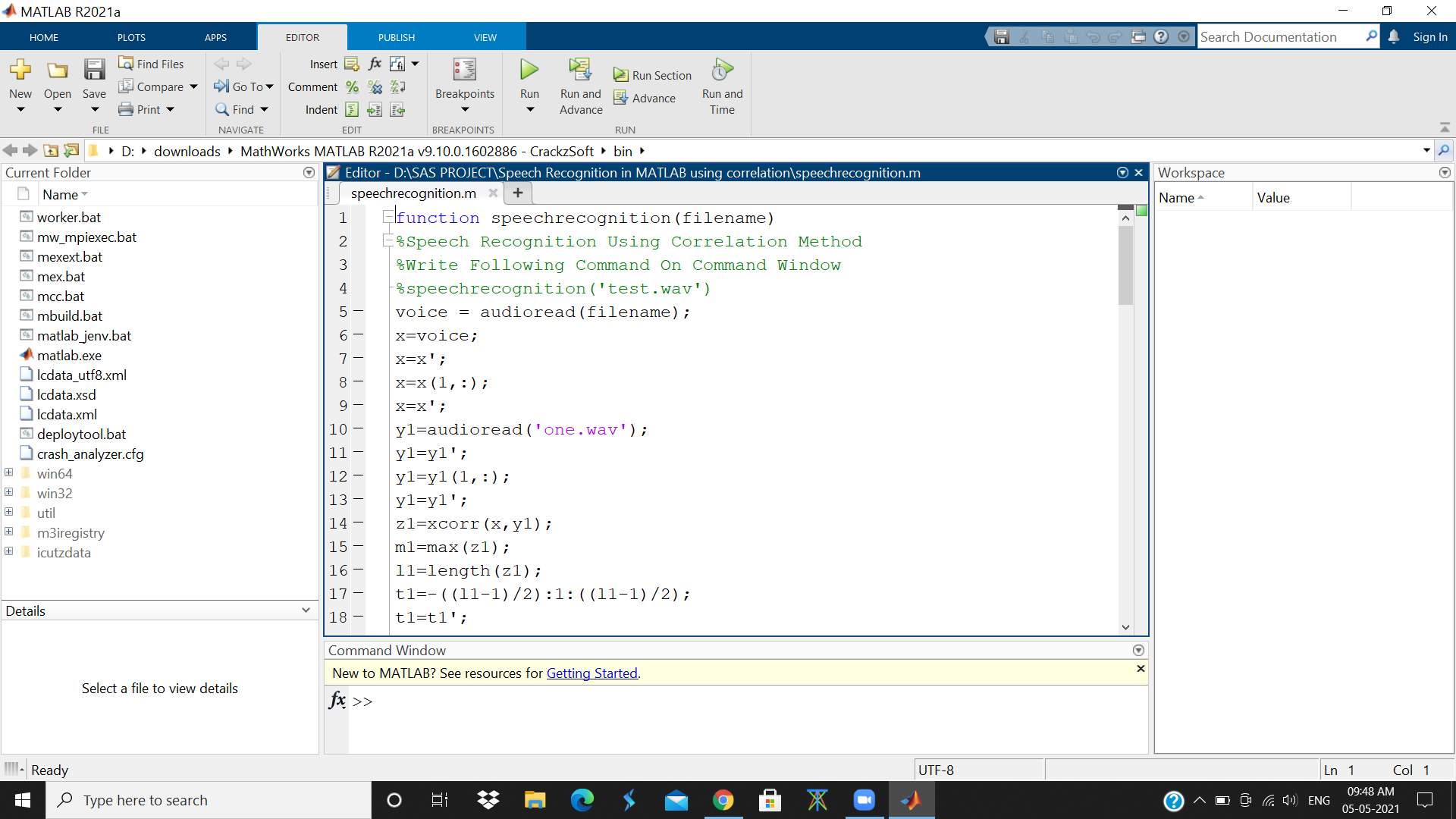
soundsc(audioread('five.wav'),50000)

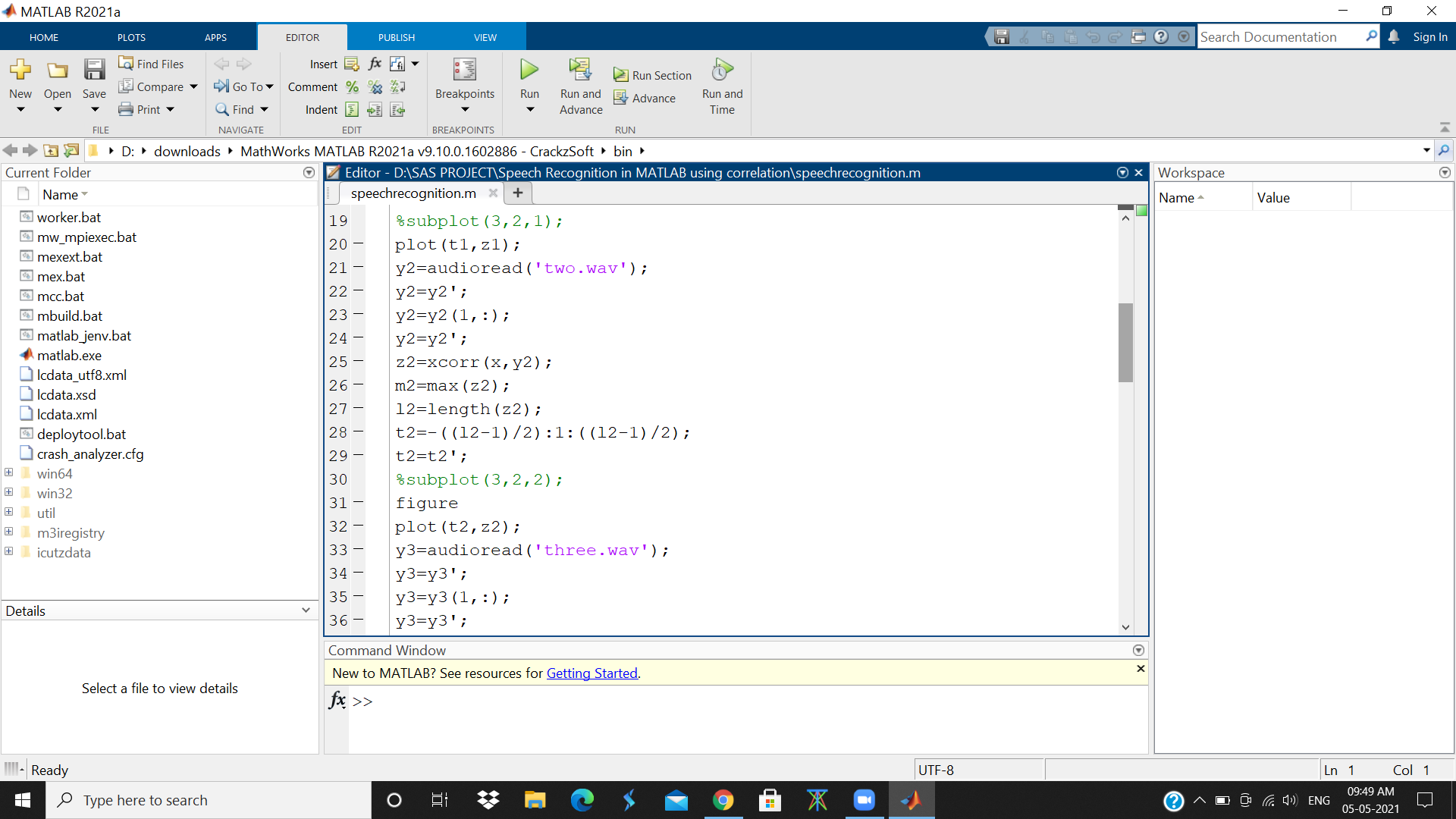
soundsc(h,50000)

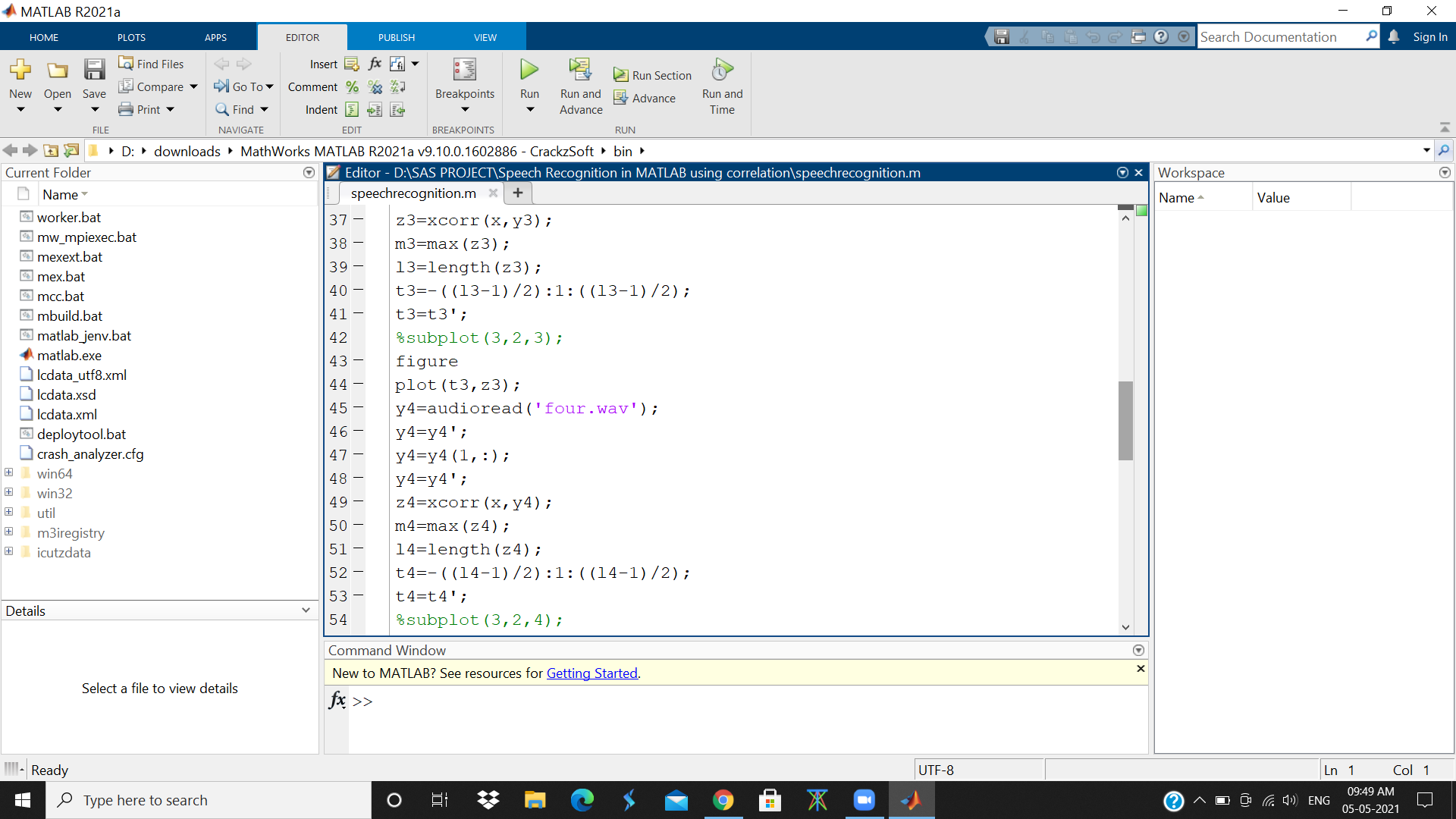
else

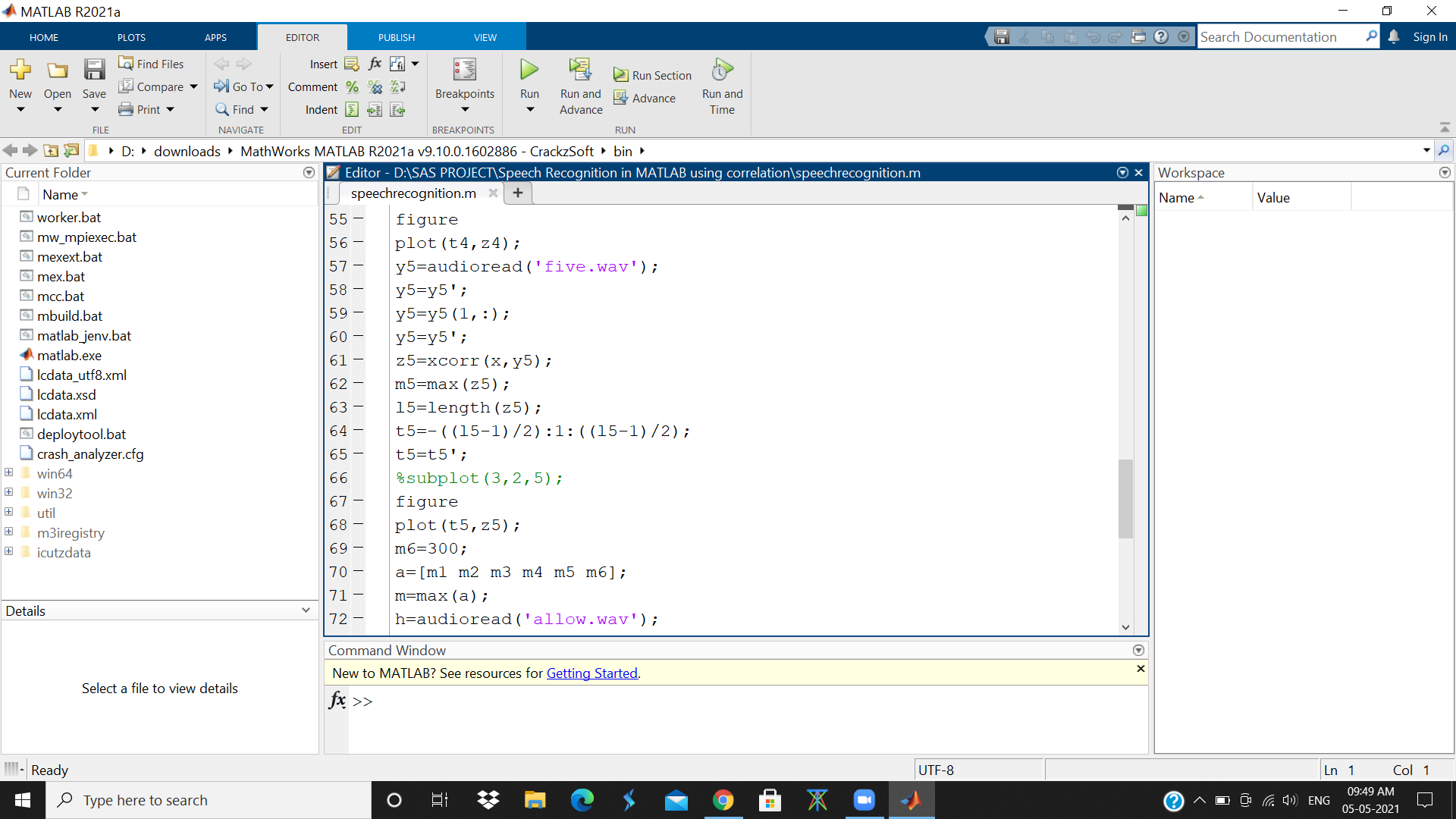
soundsc(audioread('denied.wav'),50000)

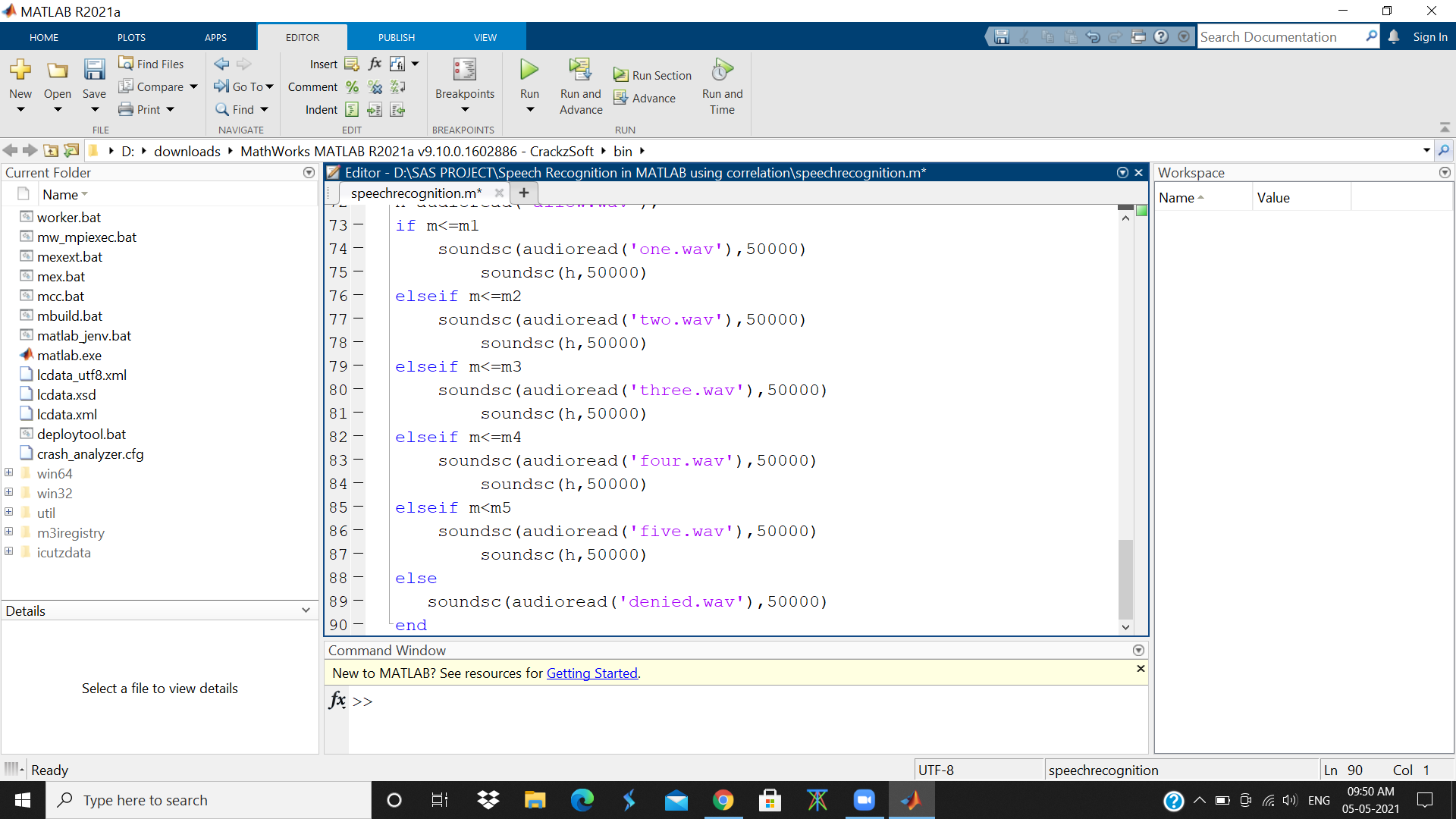
end

****

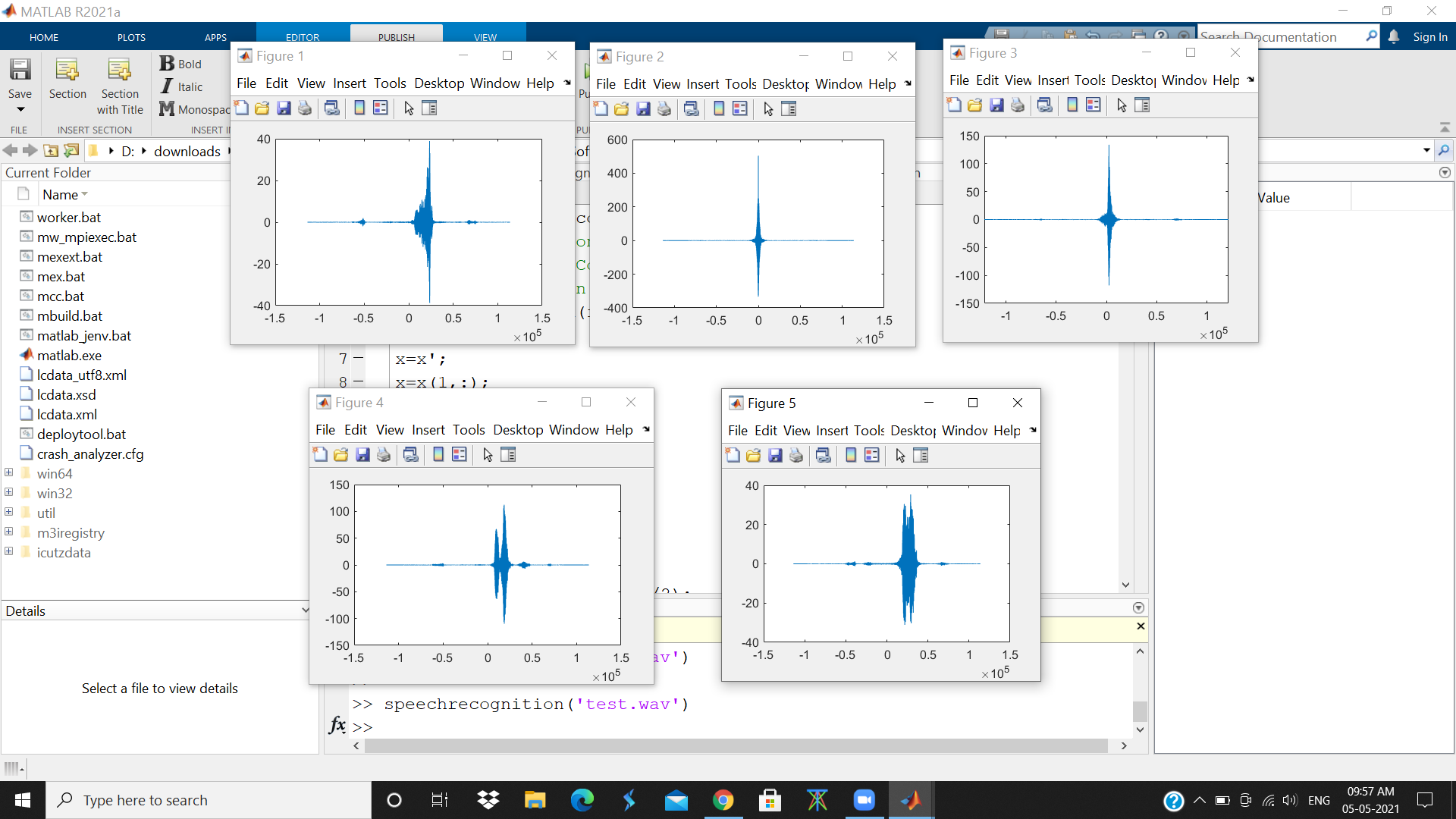
****

****

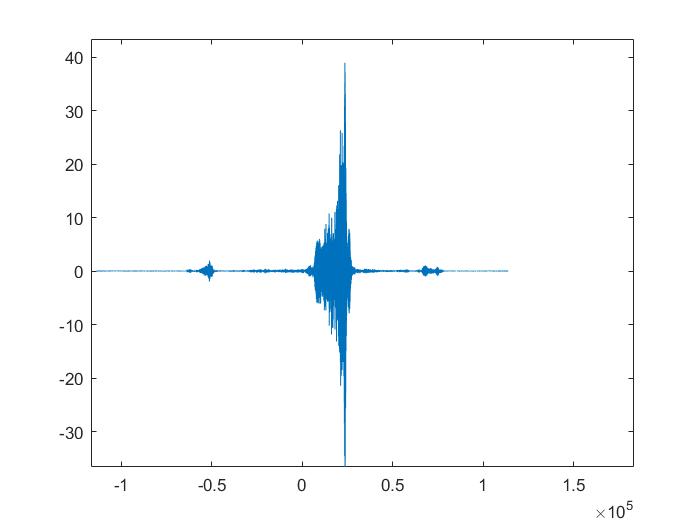
****

****

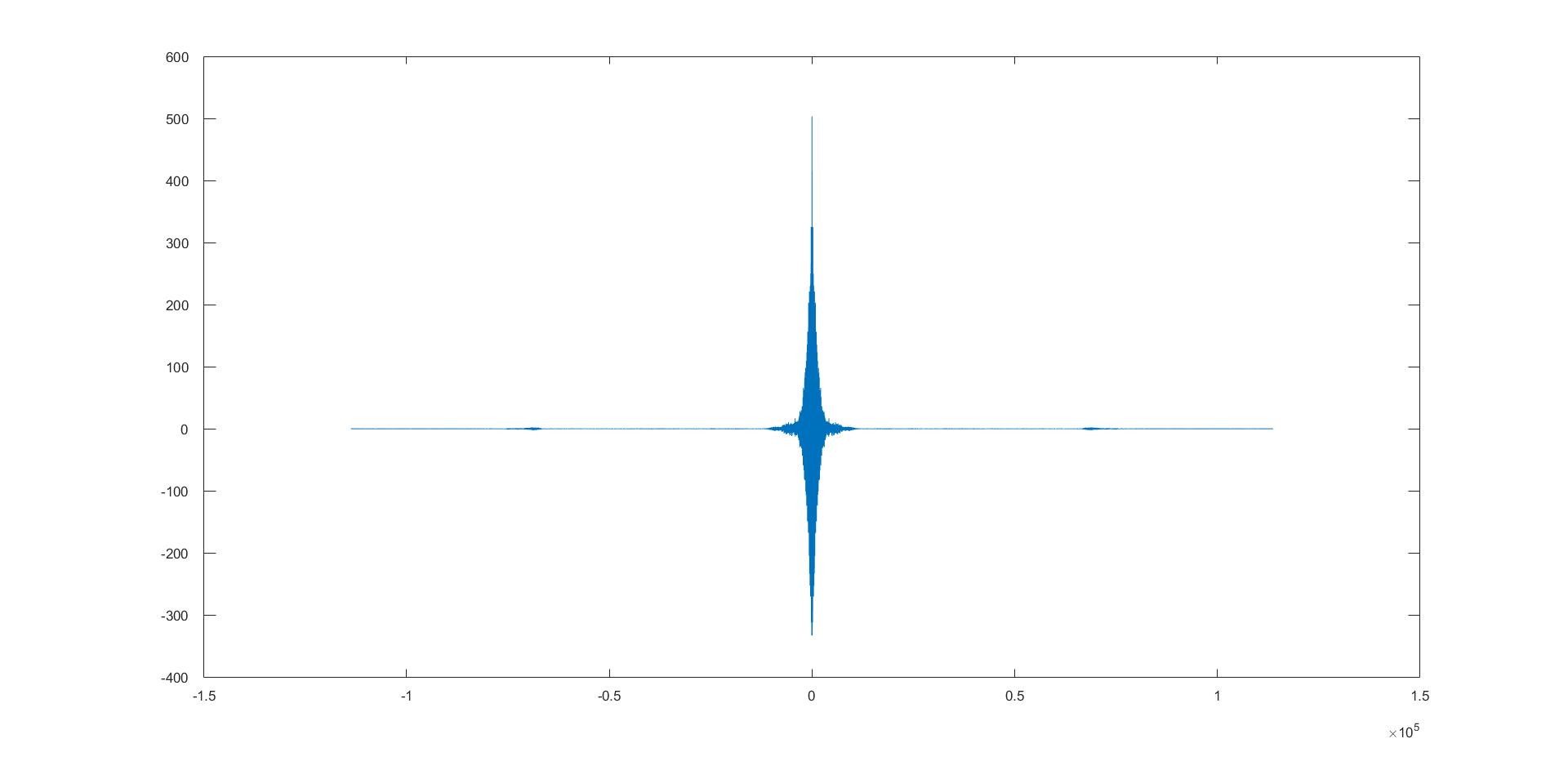
**OUTPUT:**

****

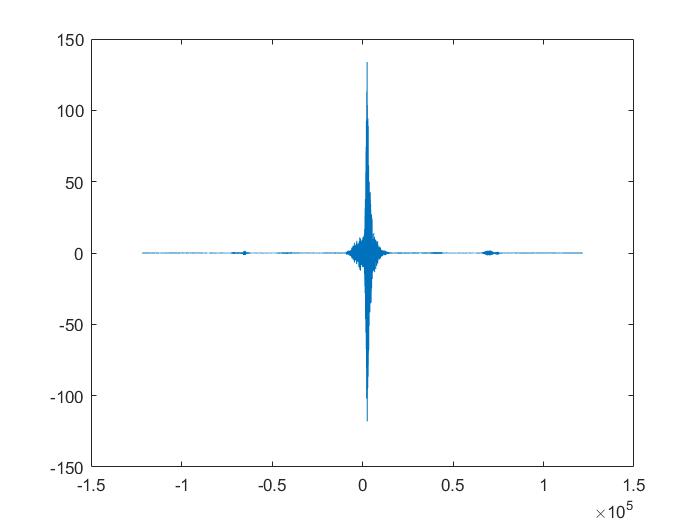
**Test.wav output**

****

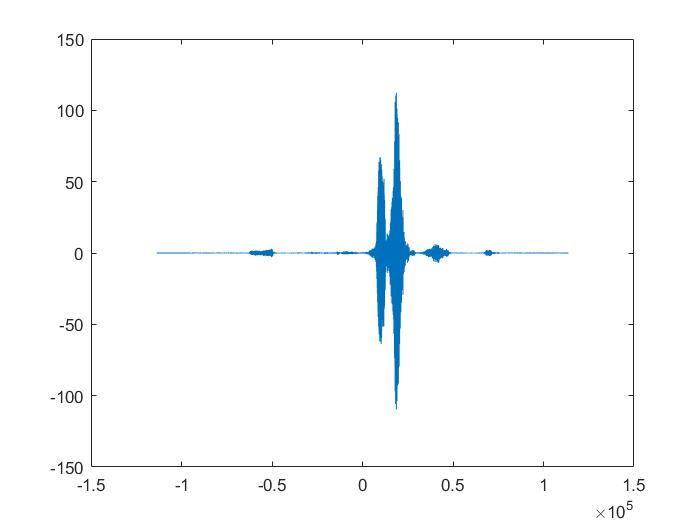
**Figure 1: Spectrum Graph**

****

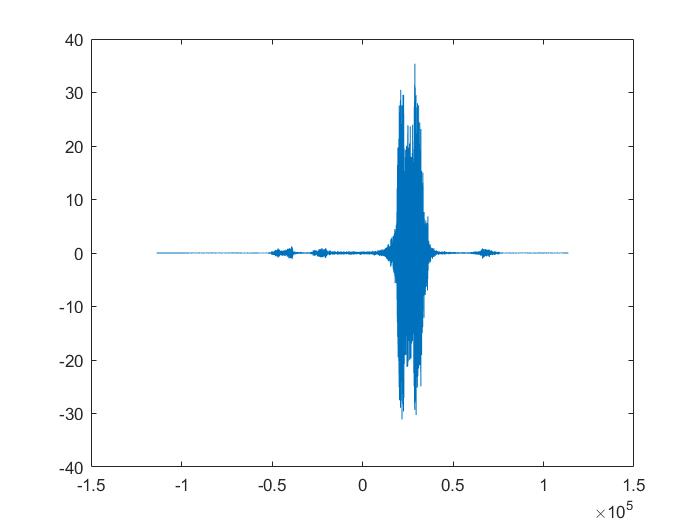
**Figure 2: Spectrum Graph**

****

**Figure 3: Spectrum Graph**

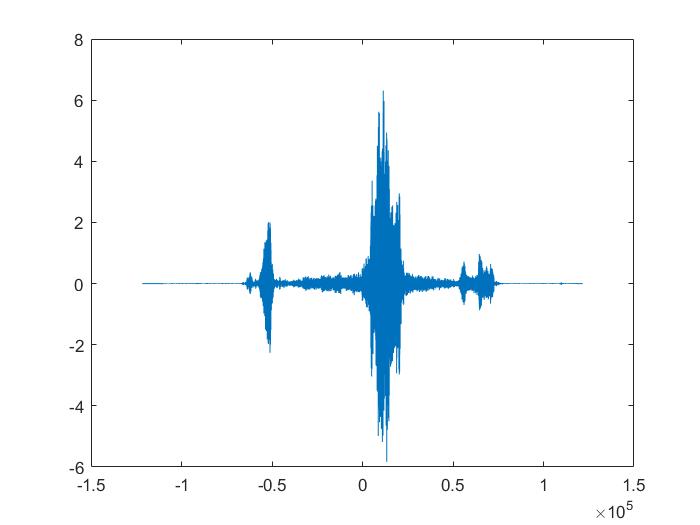
****

**Figure 4: Spectrum Graph**

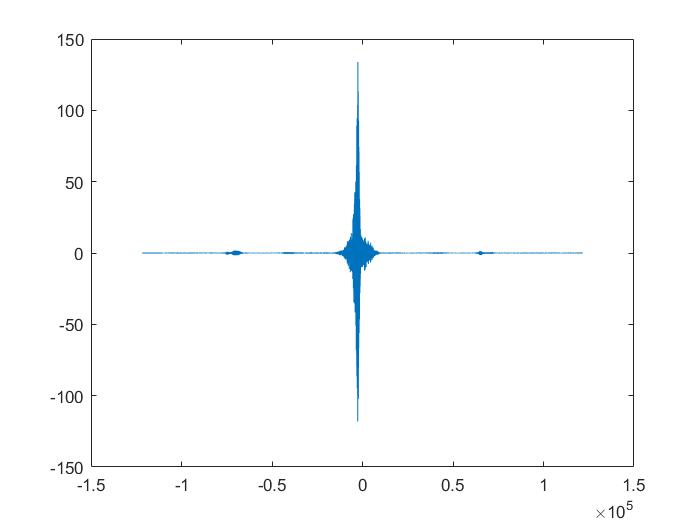
****

**Figure 5: Spectrum Graph**

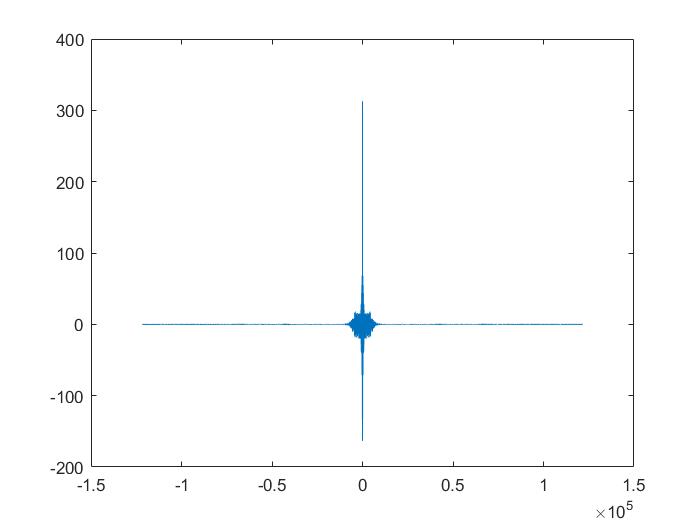
**Test2.wav output**

****

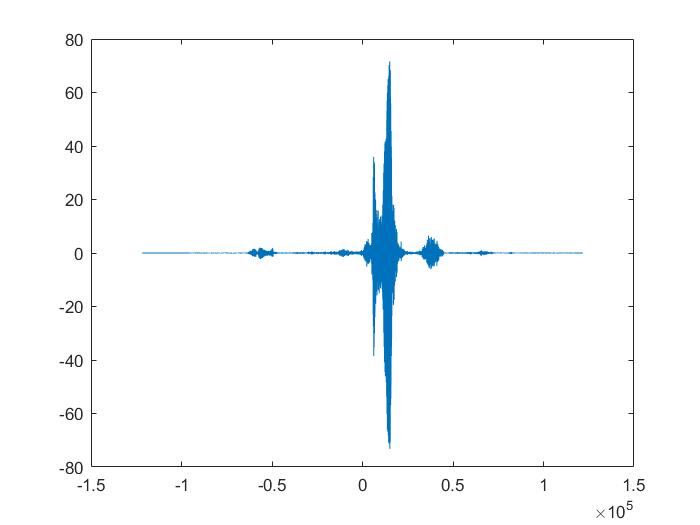
**Figure 6: Spectrum Graph**

****

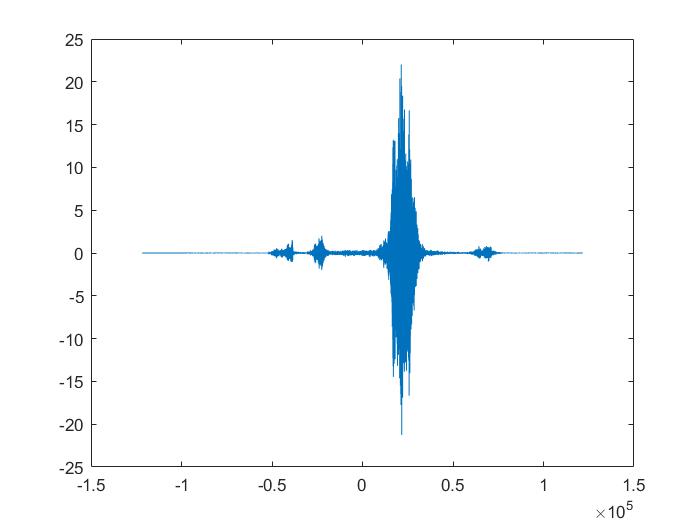
**Figure 7: Spectrum Graph**

****

**Figure 8: Spectrum Graph**

****

**Figure 9: Spectrum Graph**

****

**Figure 10 : Spectrum Graph**

**CONCLUSION :**

* 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 represent the number 2 and 3 respectively. So we ran a test for test2.wav. Now because test2.wav sounds the number 3, figure 8 is the most accurate spectrum. We repeated the step for test.wav which represents number 2. From figure 2 we can see that the graph is more accurate as test.wav sound is two.
* In this work, the five recorded wav audio files were used to demonstrate the speech recognition.
* Using the cross correlation method to find the similarities between the recorded audio files, we were able to develop a model where machines can differentiate between commands and act upon them.
* It clearly shows that machines can understand and interact with humans fluently, although they are very sensitive to noise and pronunciations.
* The above figures demonstrate a typical scenario where the machine recognizes the audio files.
* This algorithm can also be implemented using a large database in future and the results will surely be better and the efficiency will increase accordingly.