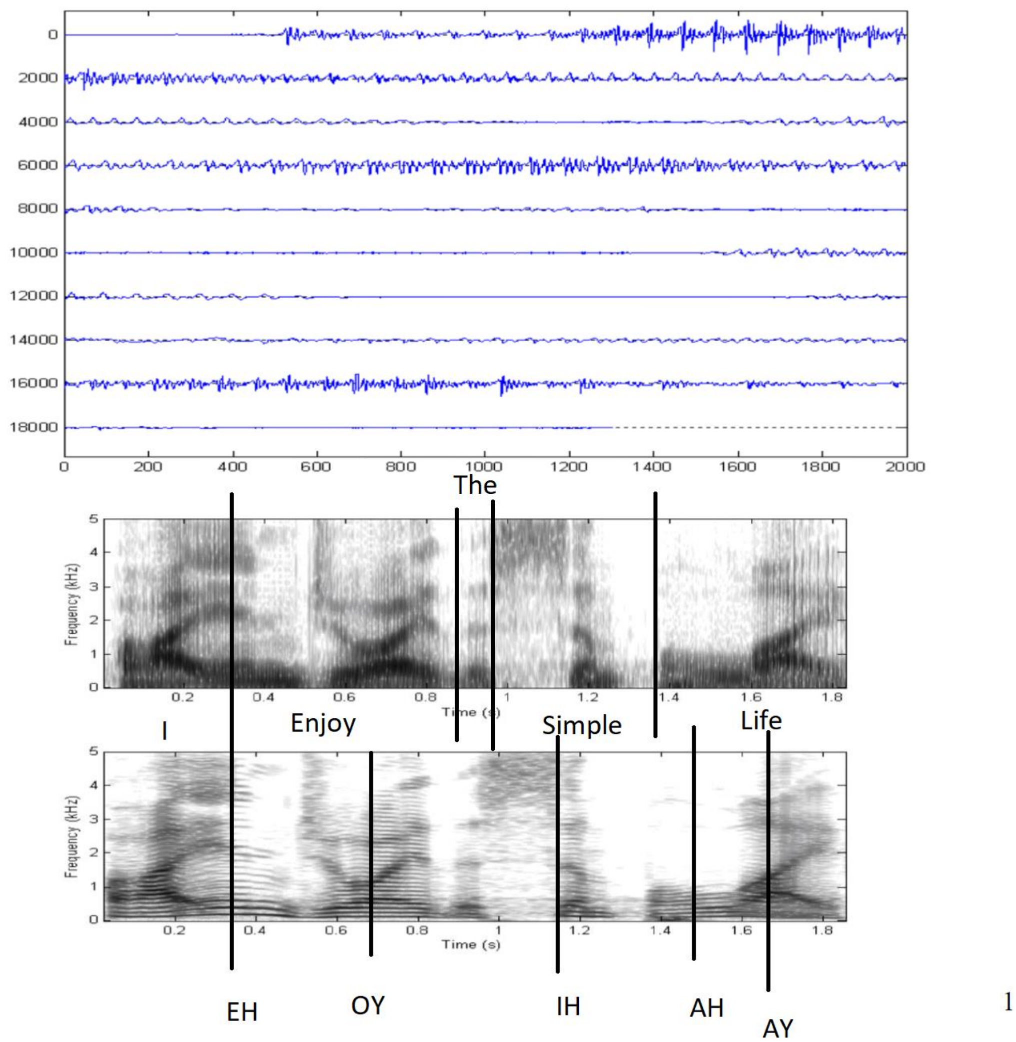
Digital Speech Processing Assignment 2 Jeet Kiran Pawani GTiD-903397407

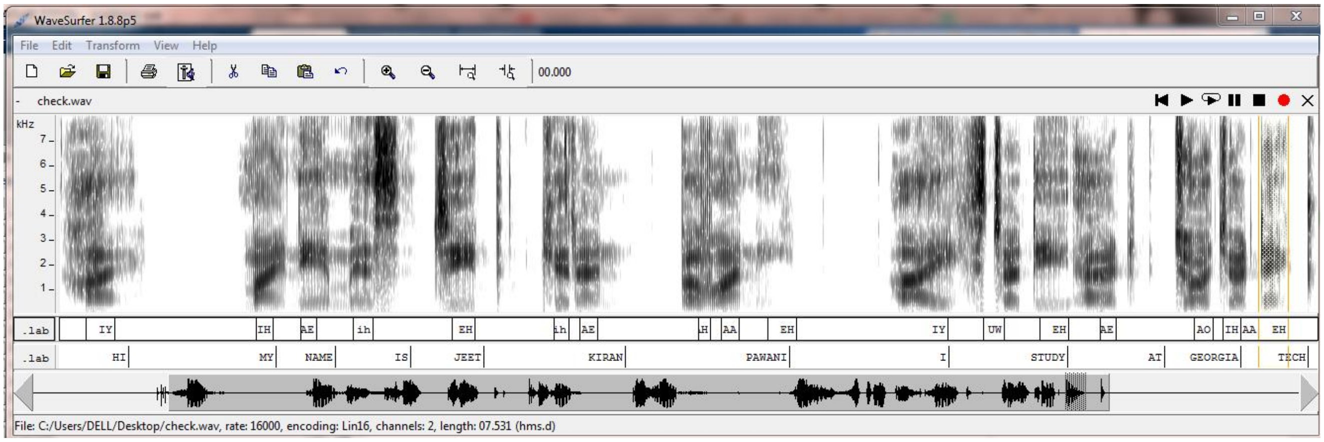
1)



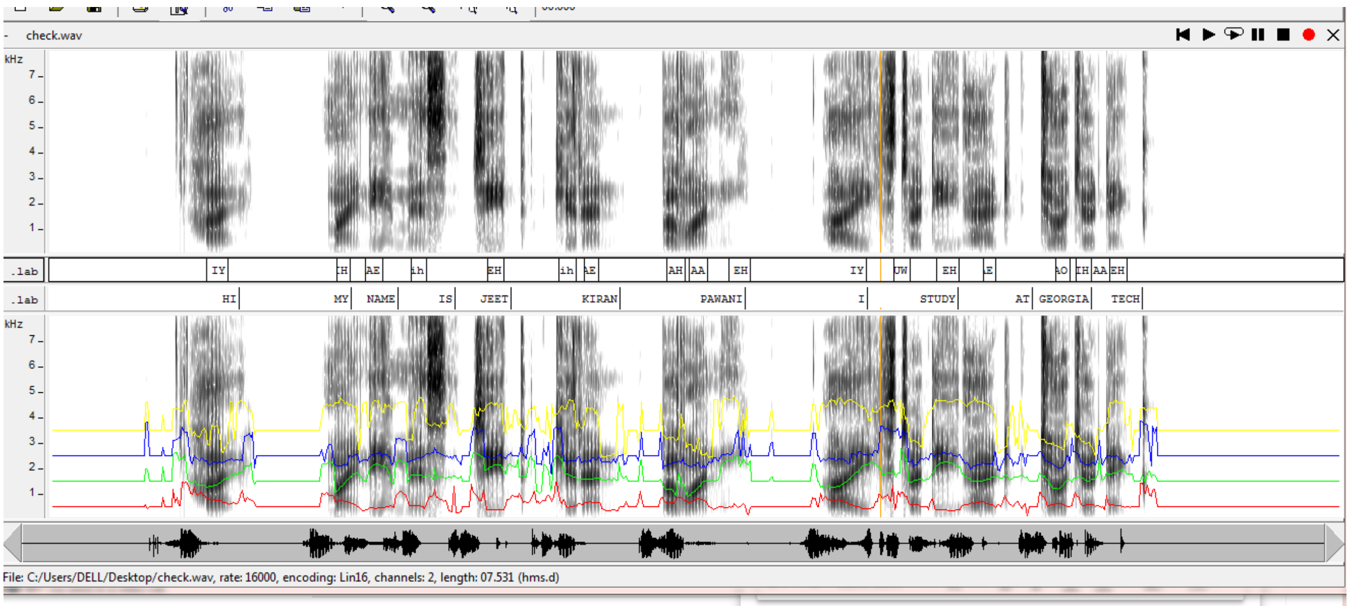
Vowel	Time(s)	Sample Number	Time spacing(approx.)	Frequency(Hz)
EH	0.36	3600	3650-3600 = 50	(1/50)*10000= 200
OY	0.63	6300	6300-6390=90	(1/90)*10000= 111.11
IH	1.18	11800	11870-11800=70	(1/70)*10000=142.8
AH	1.49	14900	14975-14900=75	(1/75)*10000=133.3
AY	1.68	16800	16890-16800=90	(1/90)*10000= 111.11
EH-OY	0.56	5600	5680-5600=80	(1/80)*10000=125
AH-AY	1.59	15900	16000-15900=100	(1/100)*10000=100

Digital Speech Processing Assignment 2 Jeet Kiran Pawani GTiD-903397407

2) a)Plot With my speech with vowels and words marked in transcription.



My speech with Formant plot as well.



b) Formant values from graph

Vowel	Part Of Word	F1	F2	F3
IY	I	320	2120	2680
UW	Study	560	1360	2320
AA	Pawani	800	1040	2320
ER	Georgia	560	1160	2120
AH	Tech	640	1200	2200

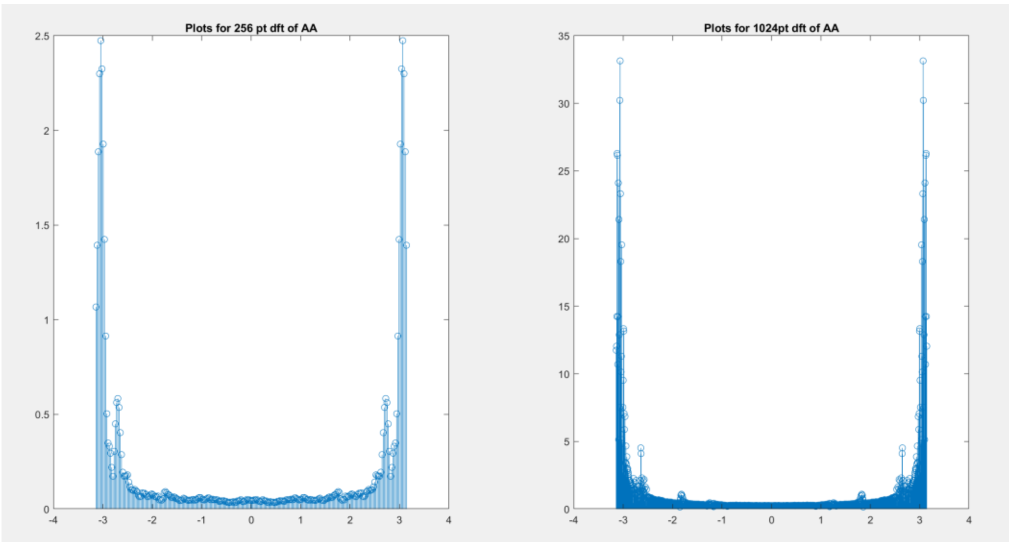
Original Formant values

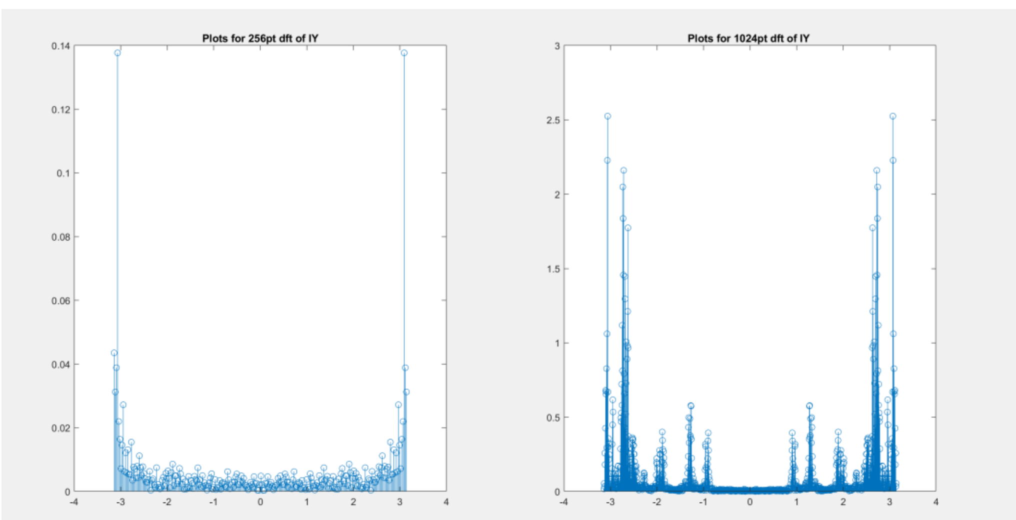
Vowel	F1	F2	F3
IY	270	2290	3010
UW	300	870	2240
AA	730	1090	2440
ER	490	870	2240
AH	520	1190	2390

c) As we can see the above 2 tables there are minor differences in the formant frequency values observed from my recording. But they are close to each other as well. The differences come up due to differences in speaking person, male/female voice as they speak at different sounds and pitch. This might be a reason there is a frequency change. Some difference might also be due to accent of the speaker as well and loudness as well.

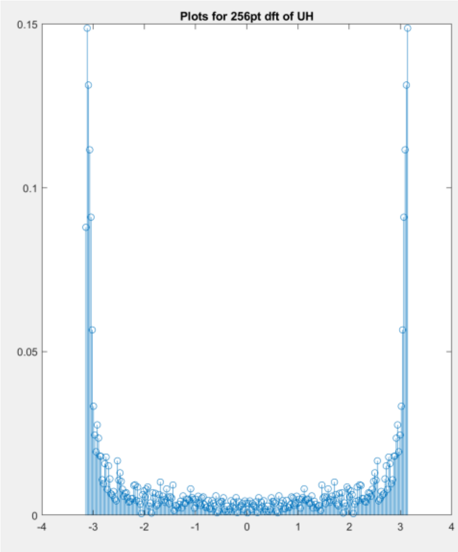
3)

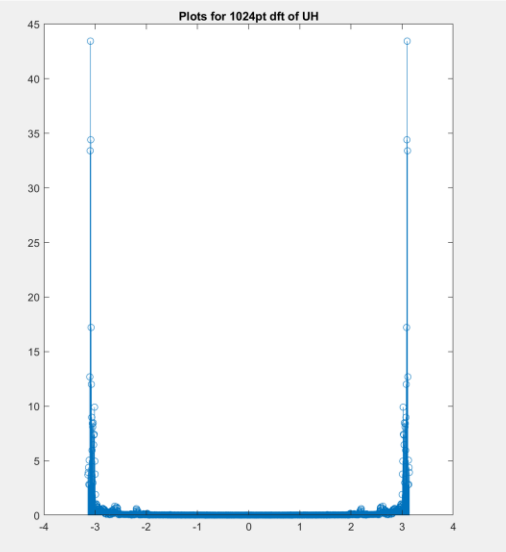
a)





Digital Speech Processing Assignment 2 Jeet Kiran Pawani GTiD-903397407



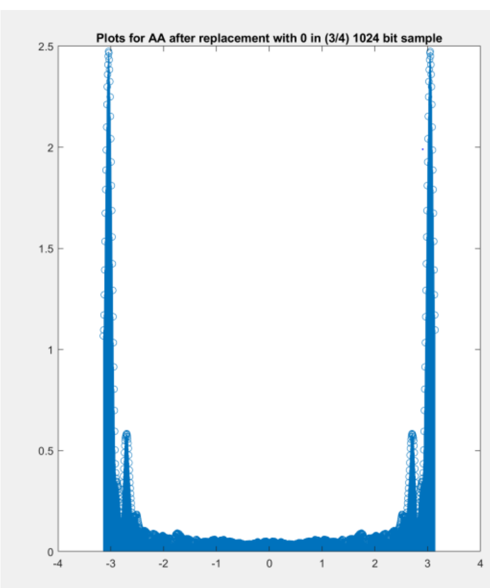


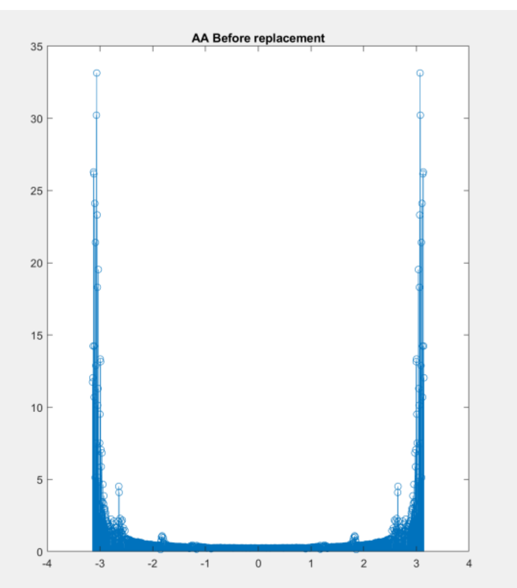
b) The differences in the plots are very evident. In case of 1024 pt DFT, there is a higher time resolution, and less frequency resolution. The frequency stems are very closely packed to each other implies more transients in the graph. Where as in case of 256 point dft, there is a lower time resolution, the frequency responses are seen clearly means high frequency resolution. The sampling frequency here is higher and thus signal peaks are clear observed.

Frequency resolution in case of 256 pt dft is (16000/256) is 62.5 and in case of 1024 pt dft is (16000/1024) is 15.625.

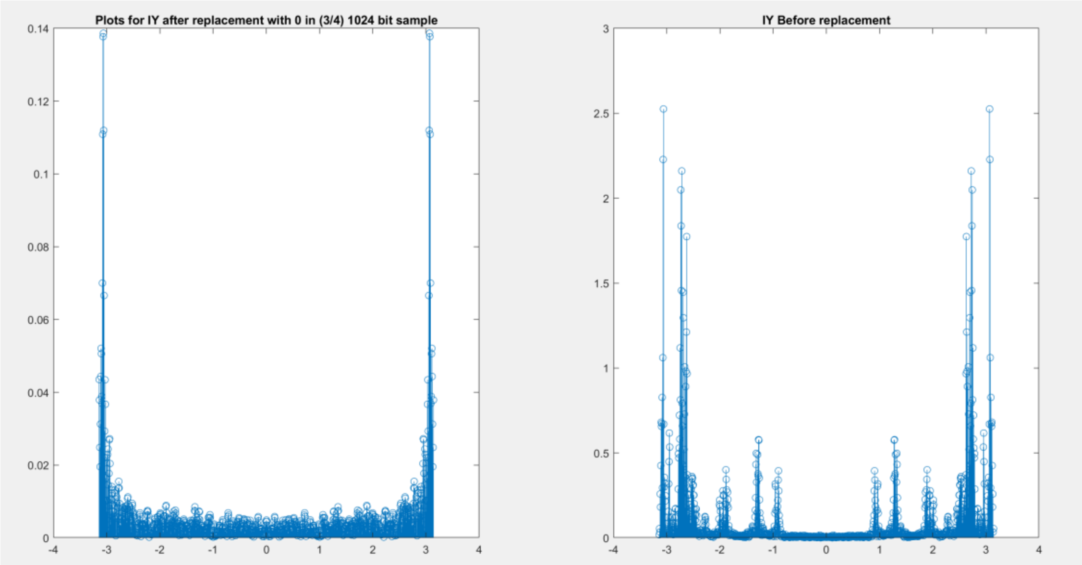
c) Frequency resolution in case of 256 pt dft is (16000/250) is 62.5HZ per sample and in case of 1024 pt dft is (16000/1024) is 15.625HZ per sample.

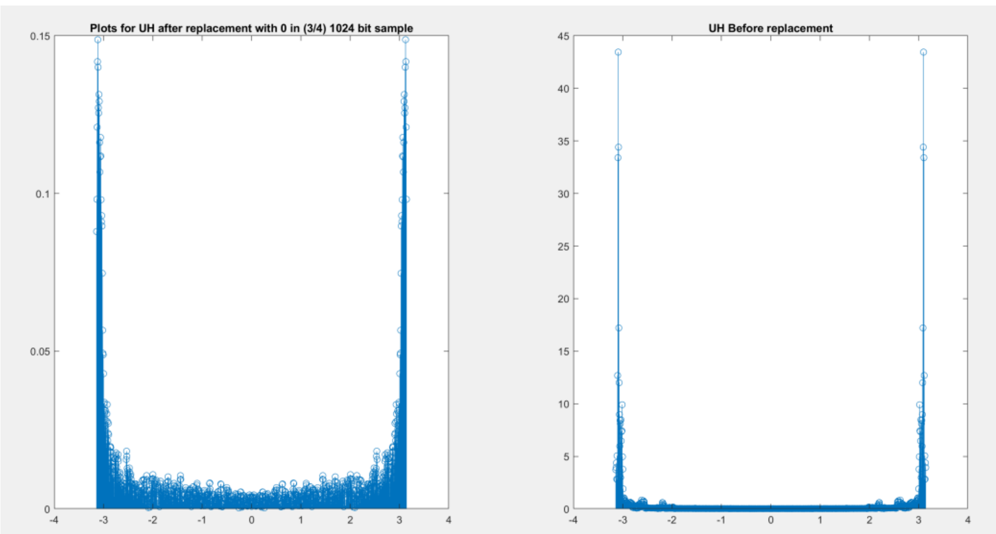
d)





Digital Speech Processing Assignment 2 Jeet Kiran Pawani GTiD-903397407

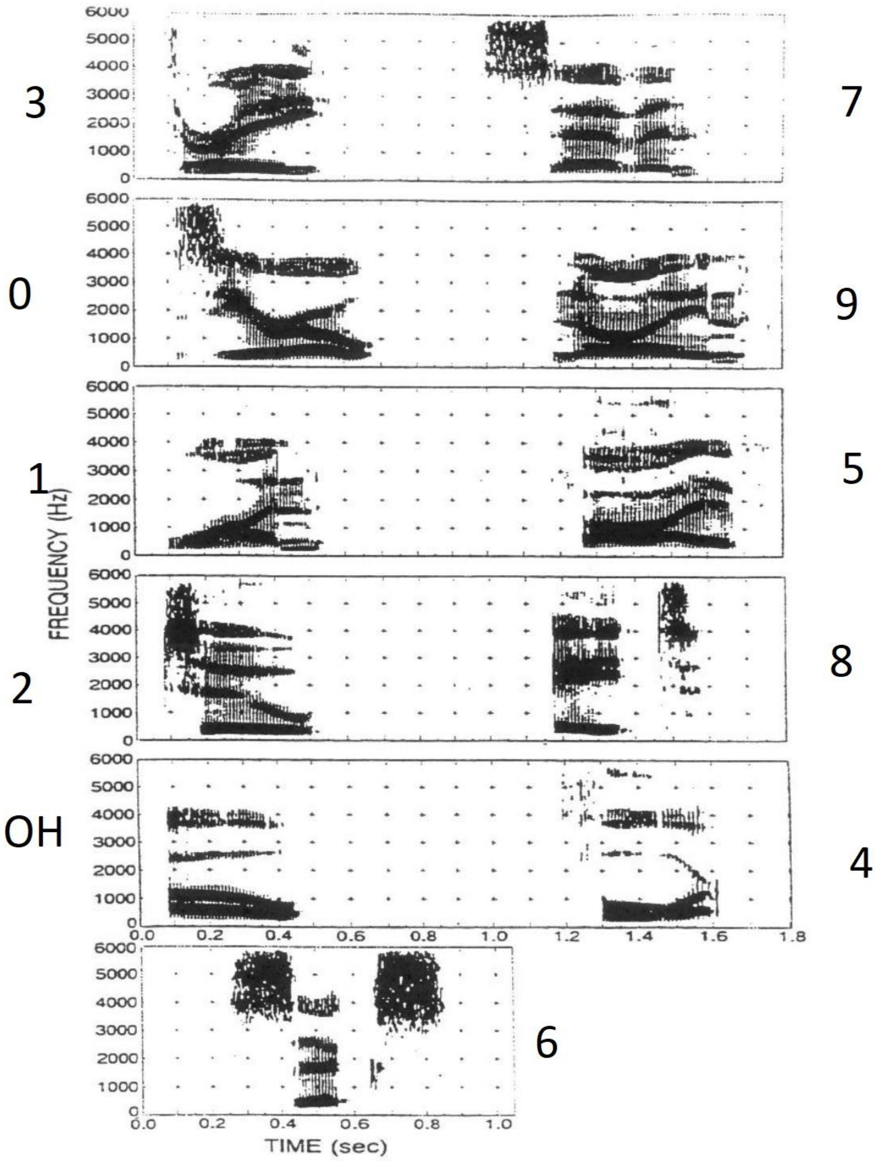




We can see in the above plots that the frequency resolution has reduced significantly, as we have zeroed out 3/4th of the sample and then performed dft on it. The spacing has increased in case of edited signal, lesser transients as 256 point signal spreads on 1024 dft. Time resolution is artificially increasing the 256 pt signal. Curve also smoothens out.

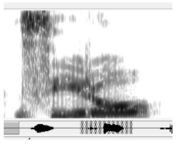
Digital Speech Processing Assignment 2 Jeet Kiran Pawani GTiD-903397407

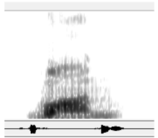
4) The appropriate digits have been written besides the images. I have also recorded my own voice of these digits so, I could verify these.



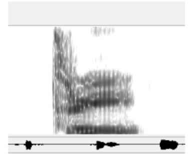
3	7
0	9
1	5
2	8
Oh	4
6	

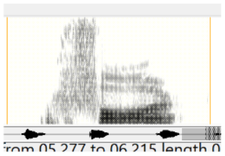
Digital Speech Processing Assignment 2 Jeet Kiran Pawani GTiD-903397407

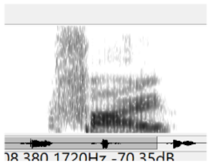

0

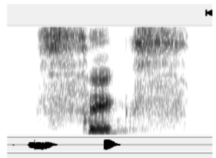

1

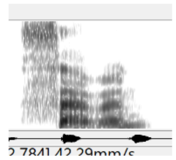

2

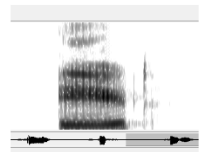

3

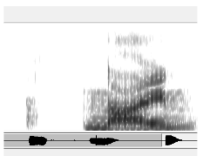

4

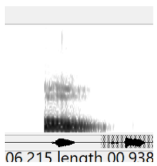

5


6


7


8


9


OH

- For guessing 0, I recorded my own voice and then found a similar spectrogram and model, where after a high pitch the dark, raises and then falls which made me guess this as 0.
- Similarly for one, there was a low pitch sound and the spectrogram looked similar.
- For 2 , there was a high pitch first followed by low sound of UH.
- For 3, the spectrogram was almost similar.
- For 4, the initial high fricative is seen and then low frequency comes up,
- For 5, its similar to 4 but in the ending of 5 , there is a higher frequency dark patch seen.
- For 6, there are 2 high pitches and in between dark patch.
- For 7, There is a gap in between where “v” sound comes.
- For 8, the graph was similar to the spectrogram given
- For 9, the high frequency in between made me guess that.
- For OH, the spectrogram itself was pretty evident.