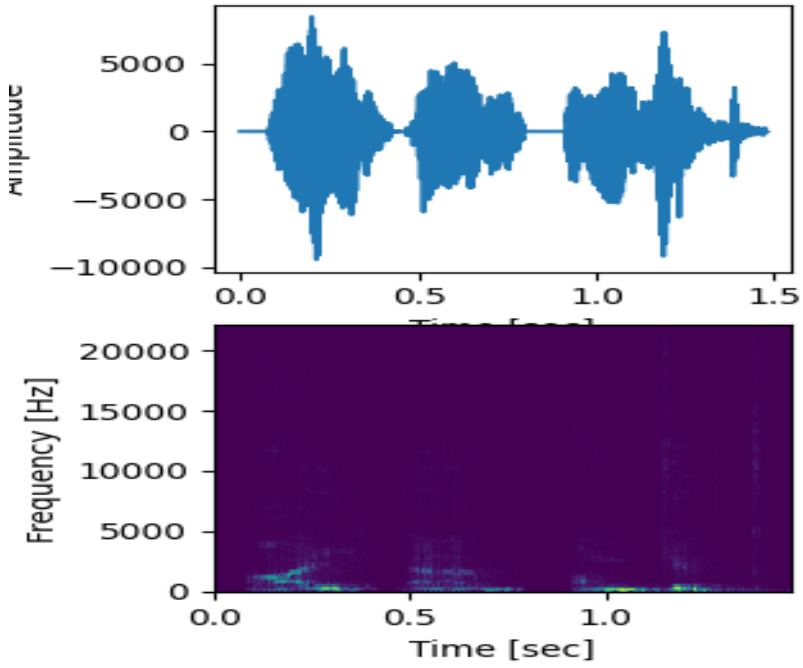
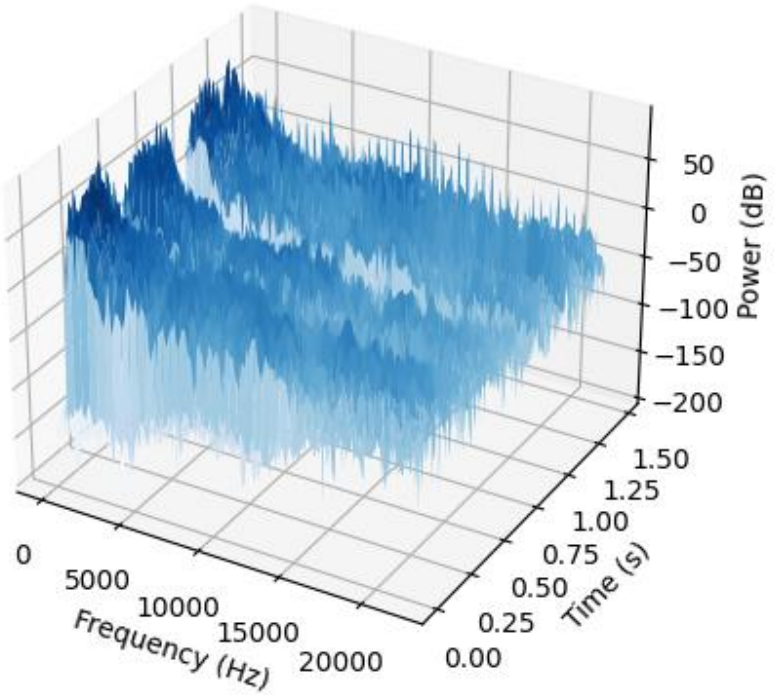
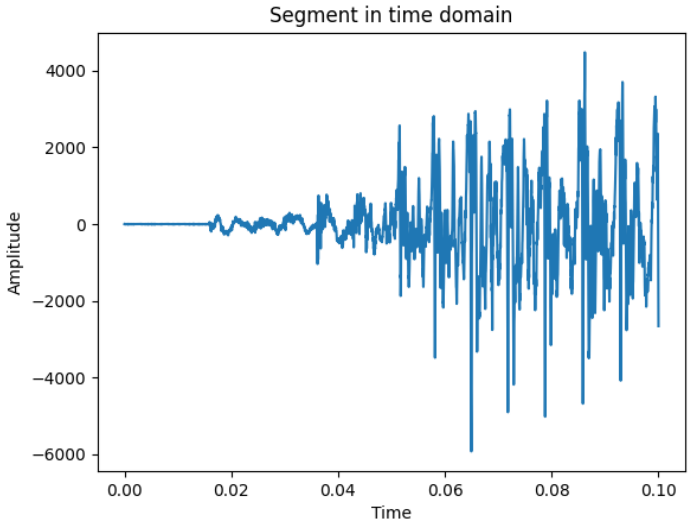
Q1. IPA transcription of the .wav file is provided **in p1_ipa.lab file.**

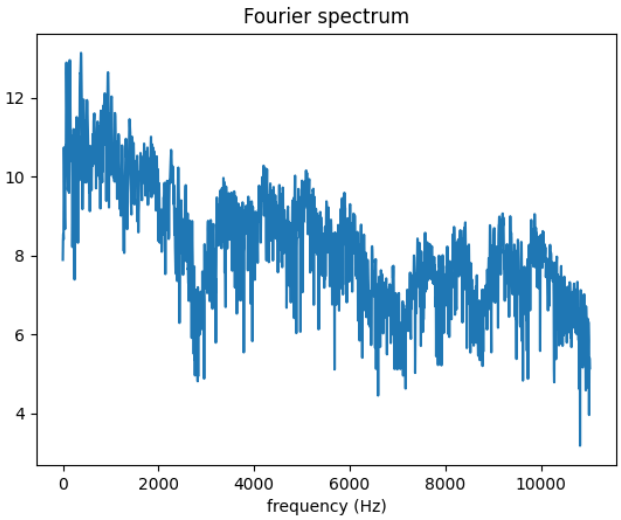


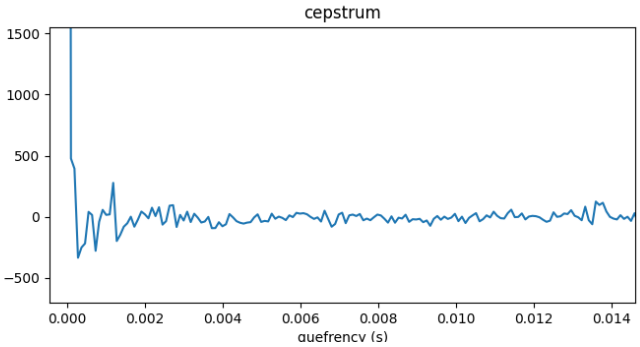


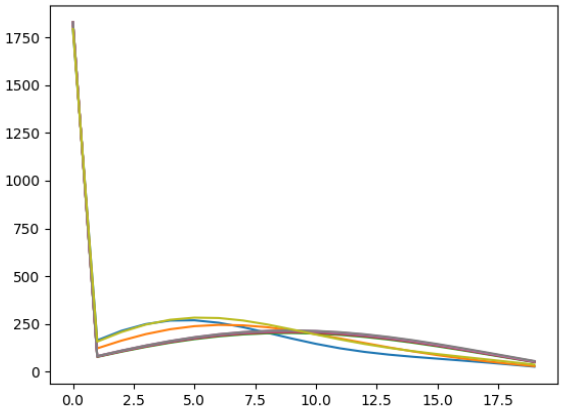
Q2. At 20ms segment was used which is plotted in time-domain . The second plot showcases the segment after being windowed by a Hamming window and then applying a DFT to it. The third plot showcases the cepstrum after undergoing subsequent log and real fft transformations. And is heavily **zoomed in.** The last plot shows the liftered signal. A sinusoidal liftering was performed with lifter value of 22 and number of mfcc chosen to be 13 as per most literature. The value is chosen to be 22 because for pitch estimation we need to perform high time liftering. Higher values help in extracting the excitation characteristics whereas lower quefrenct values extract vocal tract characteristics.

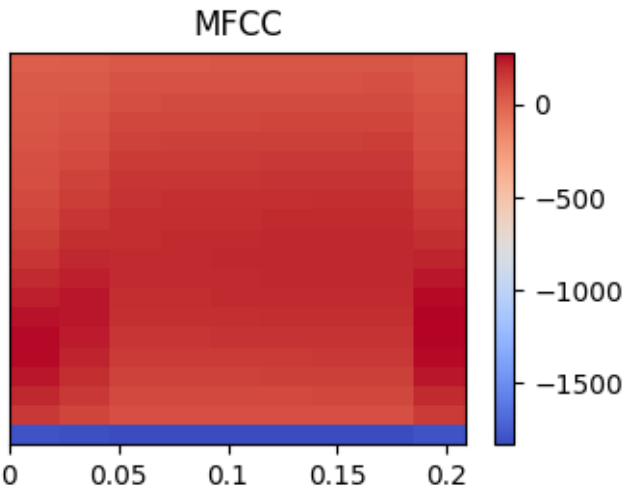
There are small pitch frequencies, however the maximum pitch frequency is observed to be 11030 Hz.





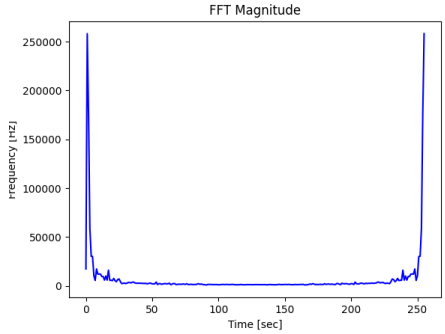

cepstrum
1500
1000
500
0
−500
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.014
quefrency (s)


1750
1500
1250
1000
750
500
250
0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

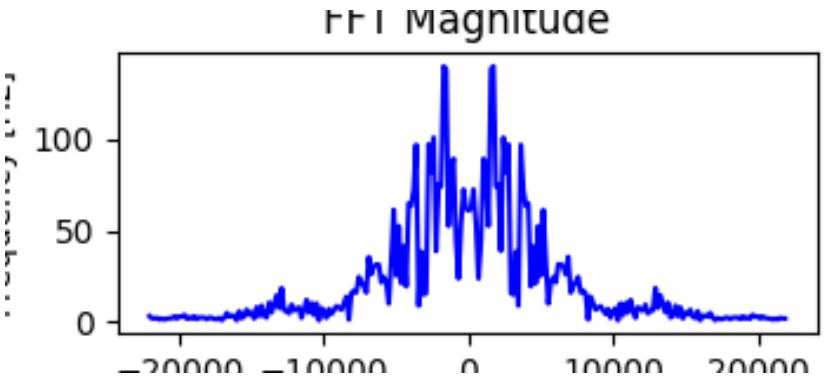

MFCC
0
−500
−1000
−1500
0
0.05
0.1
0.15
0.2

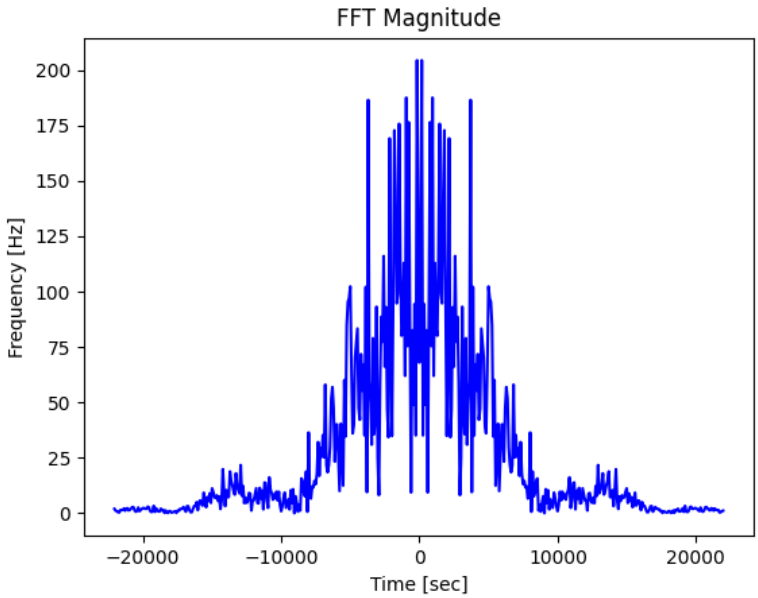
Q3.

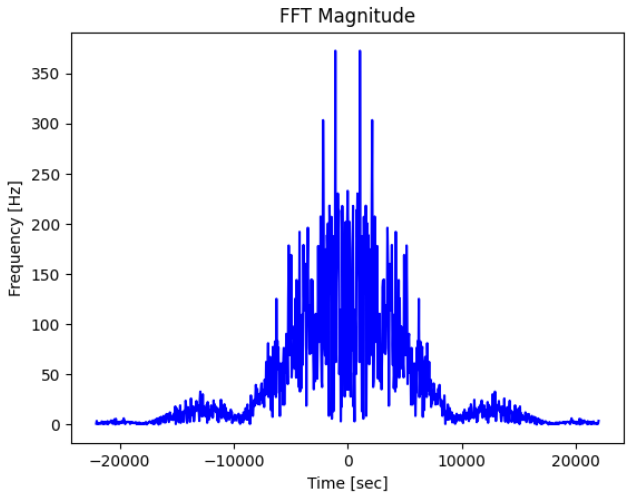
First figure shows the signal after applying FFT on **voiced segment.**



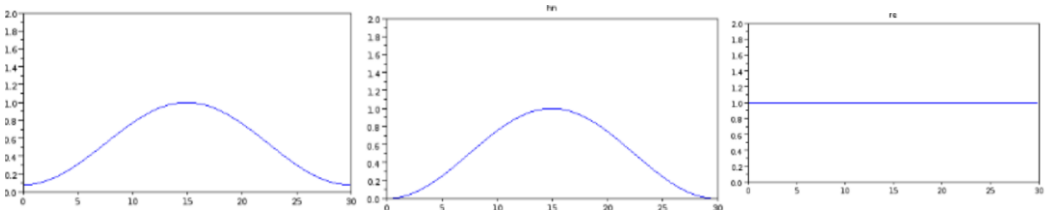
FFT from 256 ,512 to 1024. The two peaks merge into one peak towards 1024 point FFT and frequency increases. In general for both voiced and unvoiced sounds spectral resolution decreases/



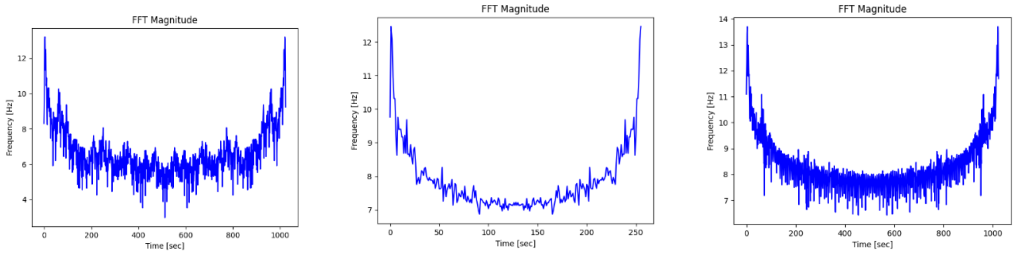




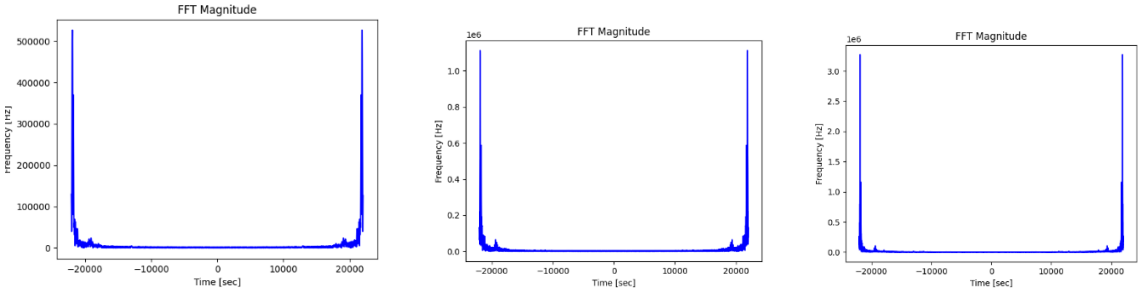
From left : Hamming, Hanning and Rectangular Windowed Signal



This was done after applying log to absolute values of FFT

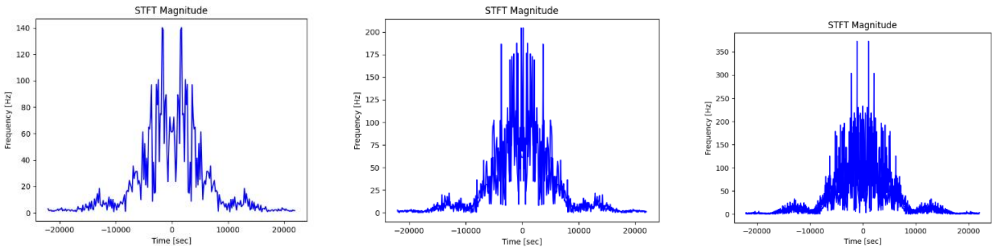


From left: FFT with 10ms, 20ms and 50ms windows

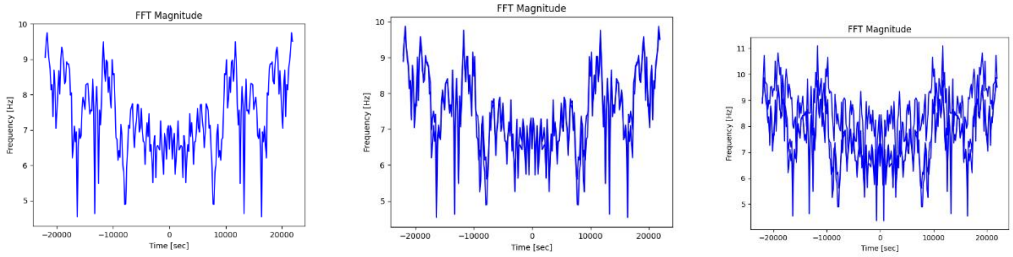


Unvoiced segment

FFT from 256 ,512 and 1024

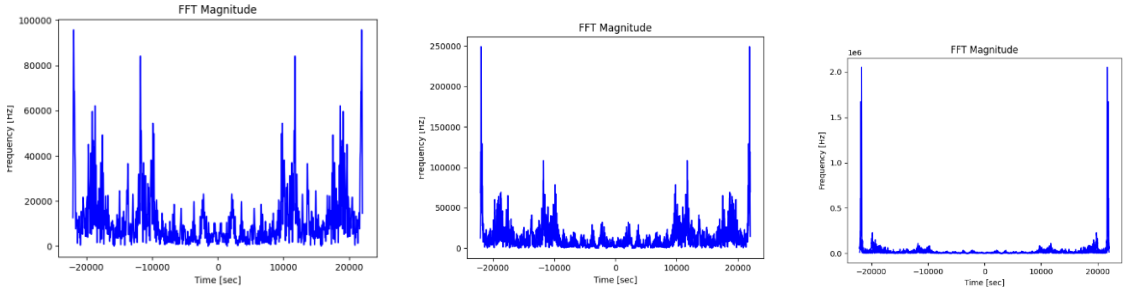


FFT from Hanning, Hamming, Rectangular



In unvoiced segments the windowed signals have significant amount of noise which is much more because of non-periodic nature of unvoiced sounds. The window is thus unable to reduce much of the leakages.

FFT 10ms,20ms , 50ms



Unlike voiced segments , in unvoiced segments there is a drastic decrease in the high frequency components on increasing the window size with FFT transformation applied.