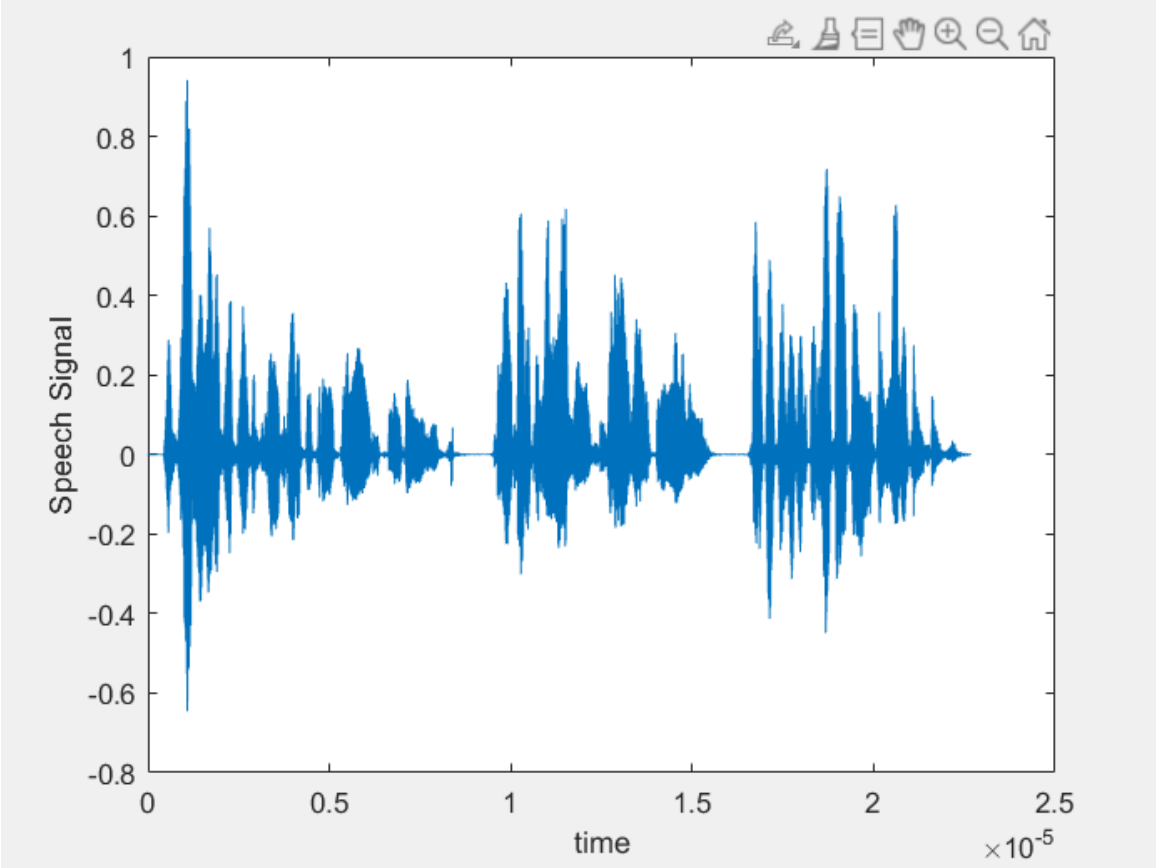
Assignment-2

Speech Signal Processing

2020102004, Anjali Singh

$Q1.$ Software used: MATLAB

(a) *Create a time domain plot.*



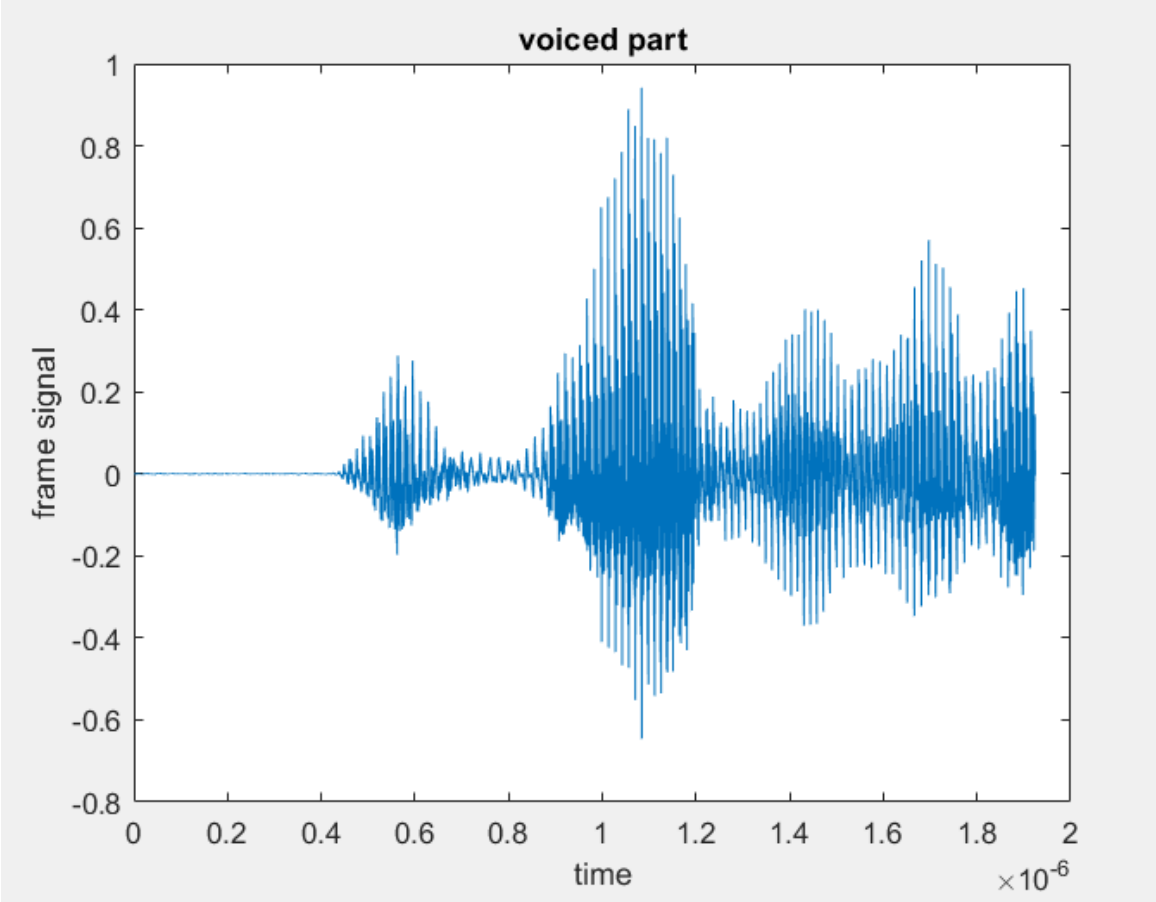
(b) *Apply framing on to the signal.*

In this part, we got 10 frames of the speech signal. I have made a function called *framing,* which is used to frame the signals.

Note: We can find any number of frames just by changing the value of *frame_duration*.

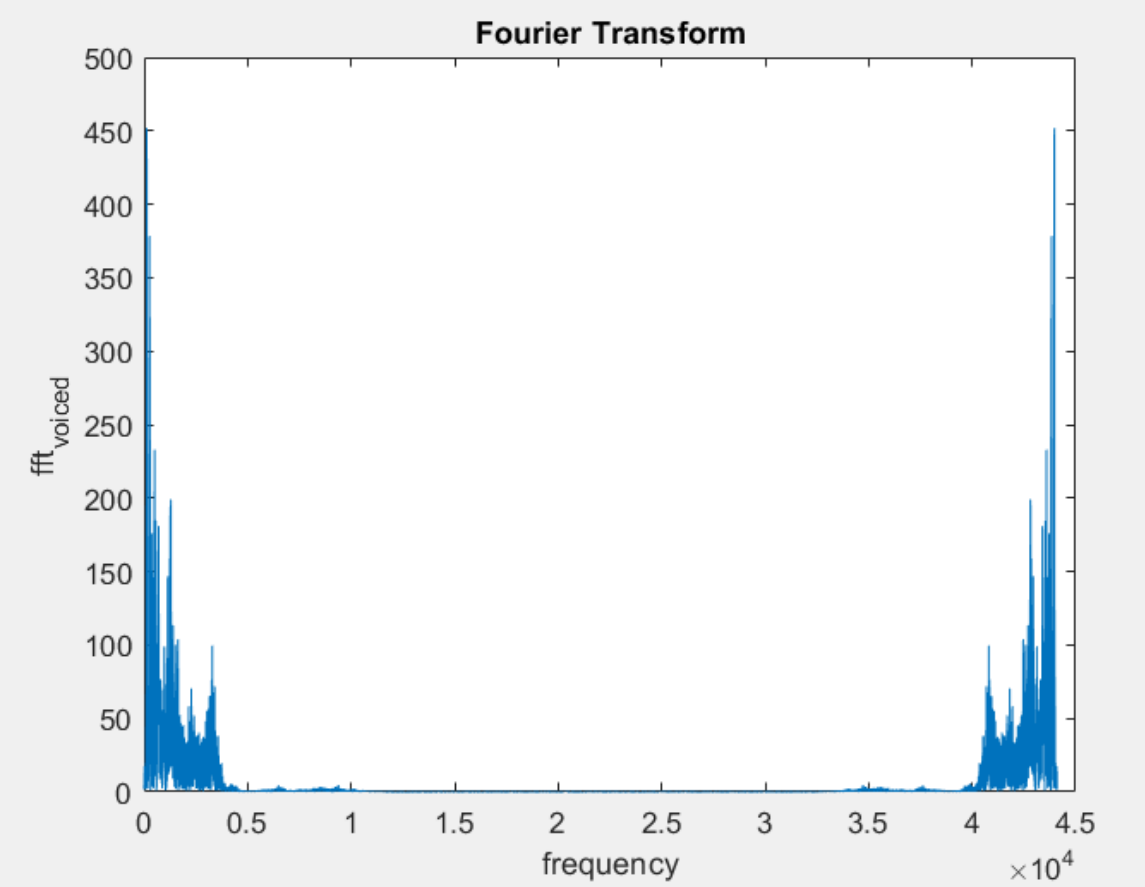
(c) *Find one voiced frame among the all the frames and plot it.*

The 1st frame, with maximum amplitude or energy is the voiced of all, calculated by finding the energy of the frames.

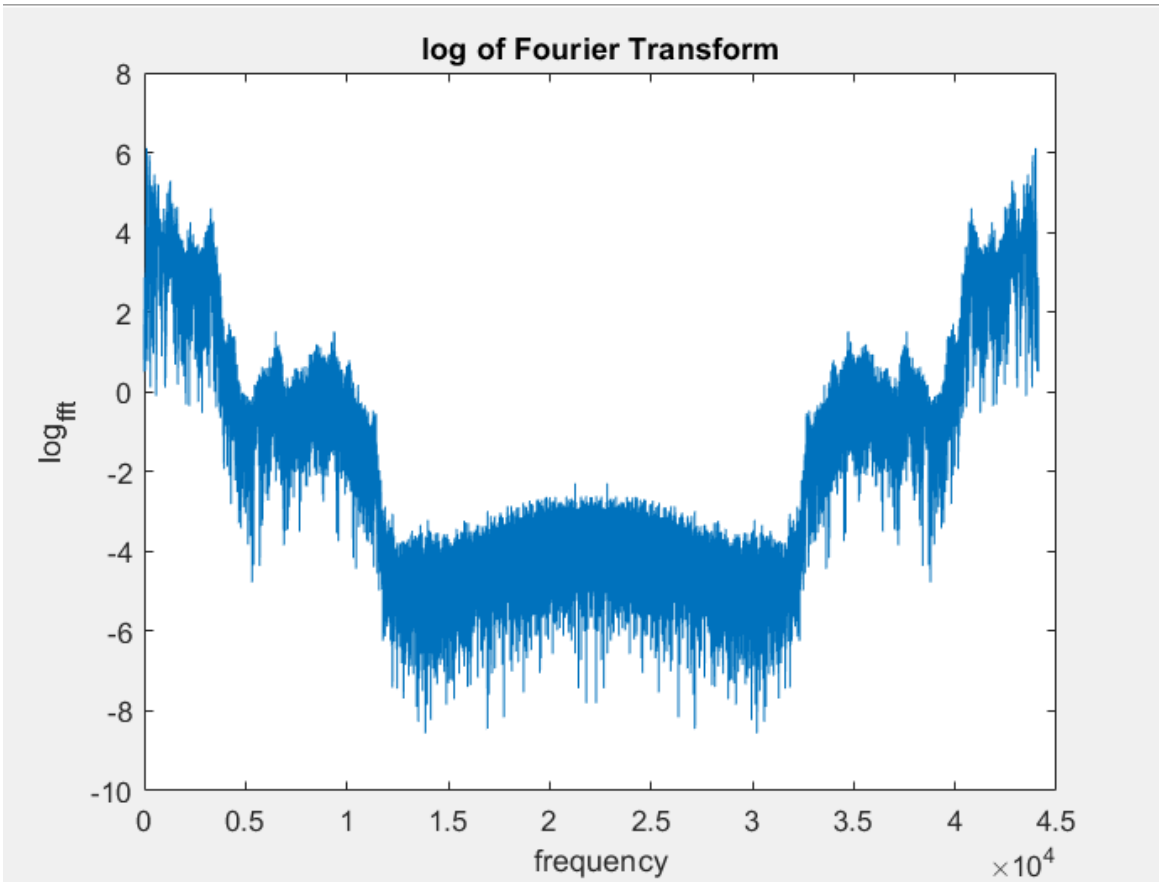


(d) *Compute Fourier transform on the voice frame which you have considered in the step (c).*

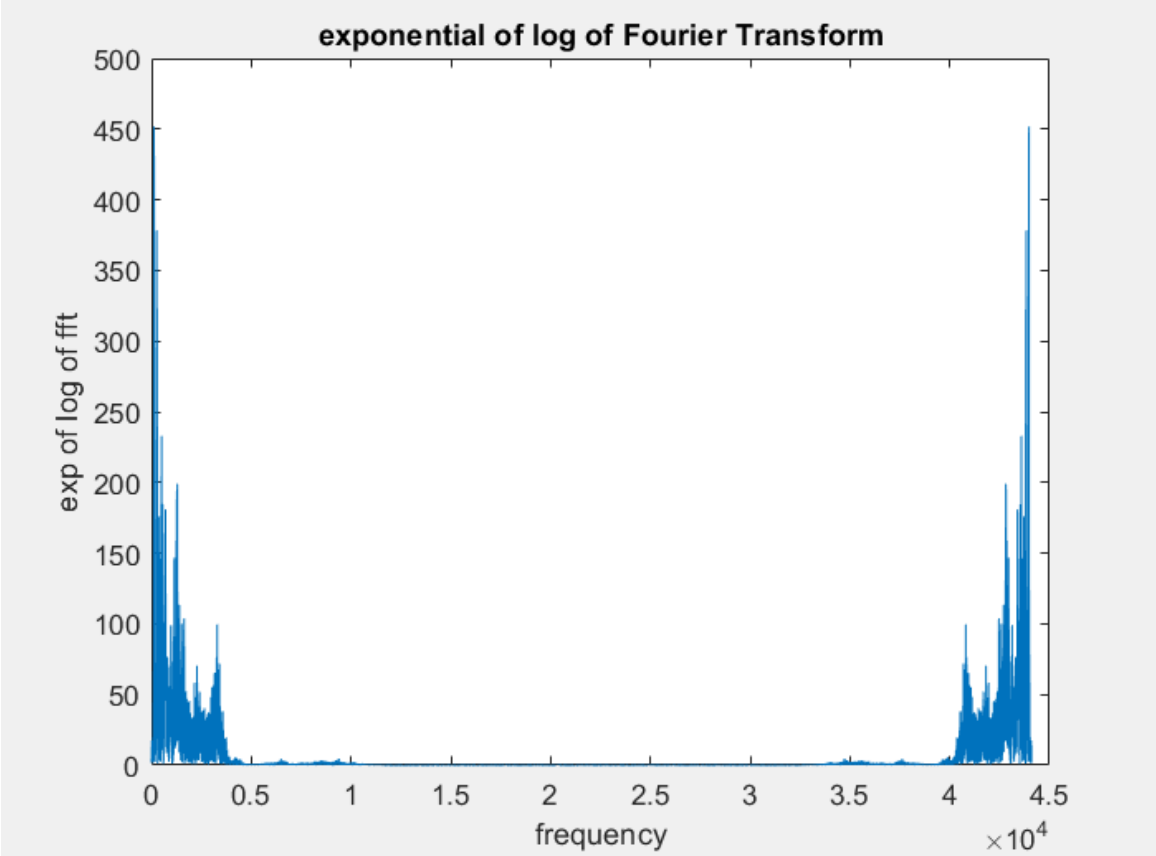
- Plot for Fourier Transform of voiced frame



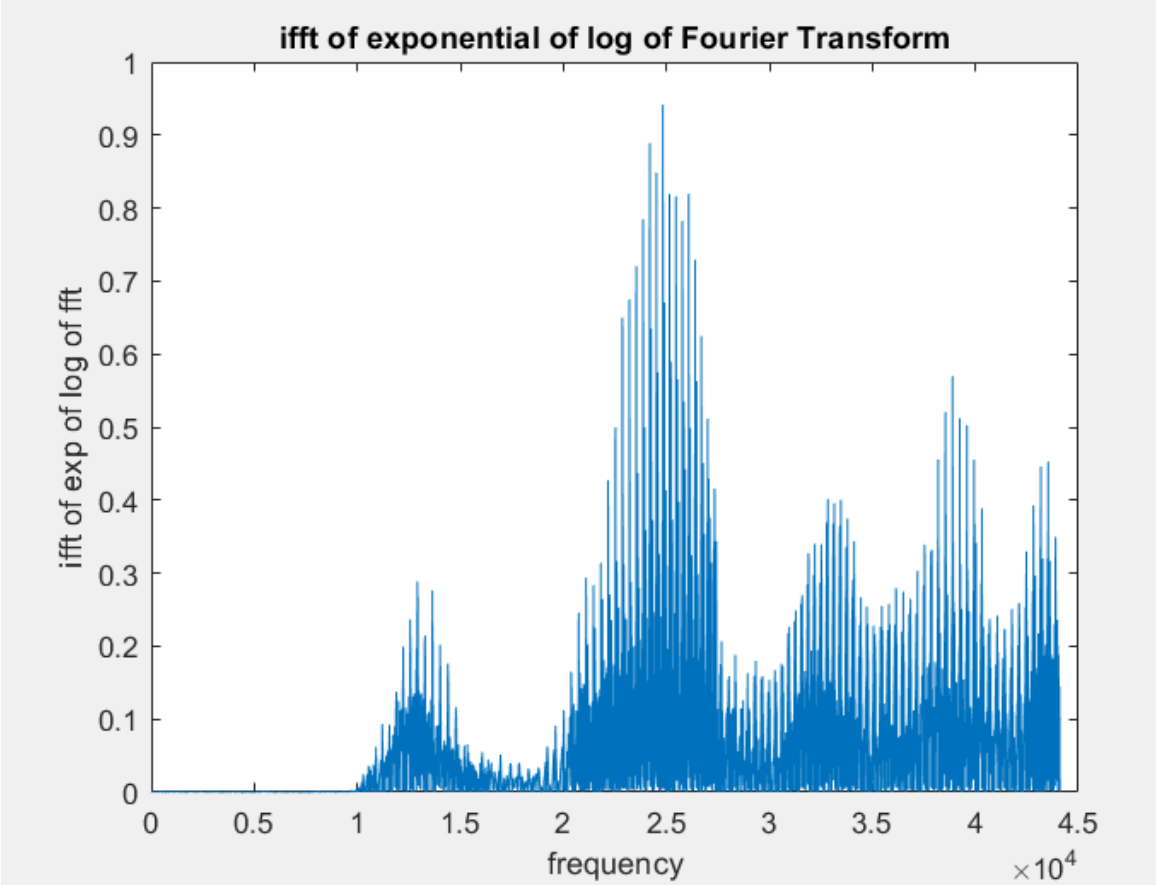
(e) *Apply log to the step (d) and plot it.*



(f) *Compute exponential operation for the step (e).*



(g) *Compute Inverse Fourier transform for the step (f) and plot it.*

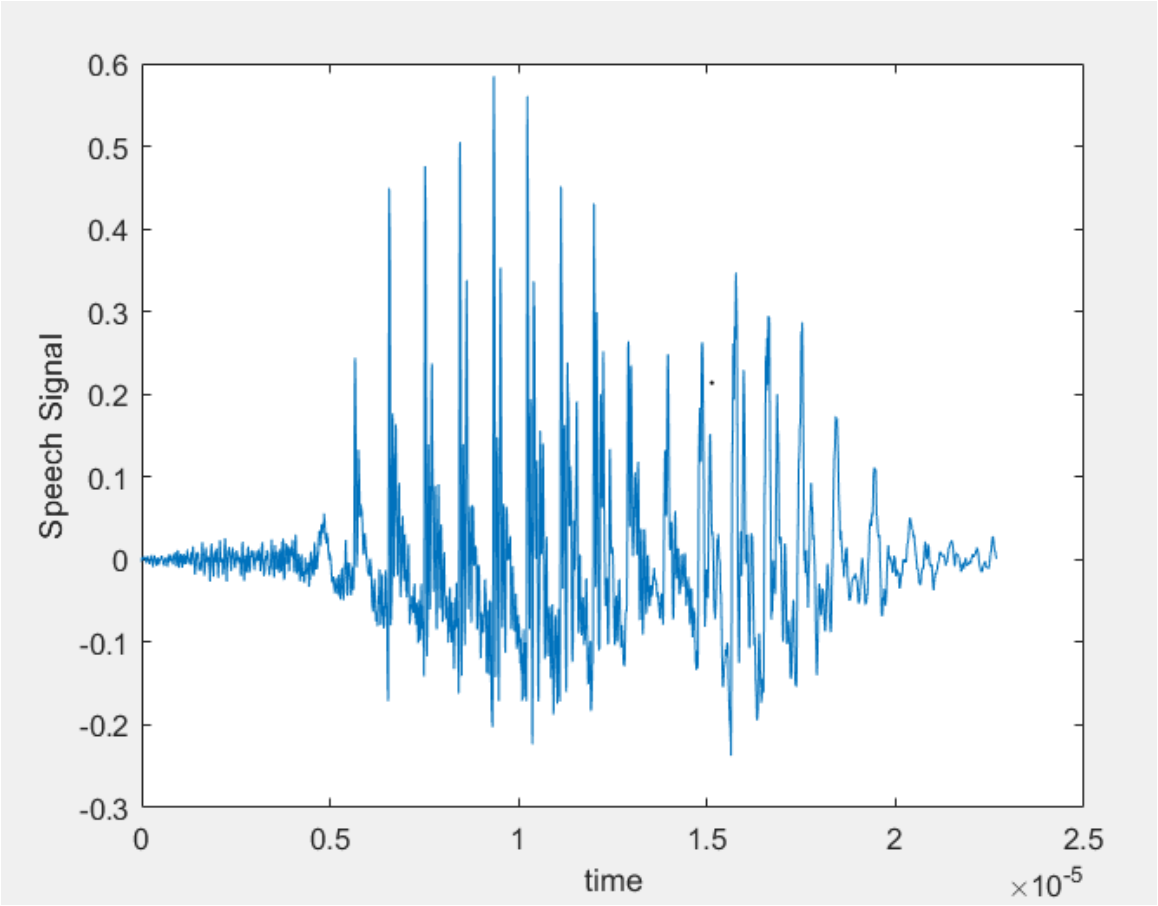


(h) *Comment on step (c) and step (g). If the outputs are same justify why?*

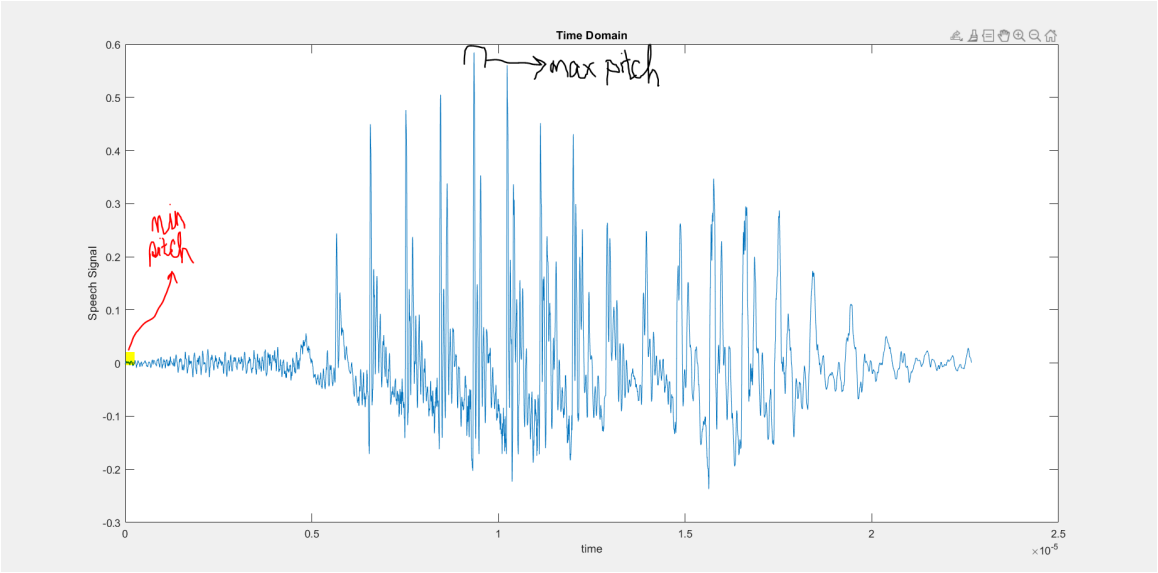
- Since, I've taken the absolute values of Fourier transform and inverse Fourier transform, because the values contained complex values too.
- On comparing the plots of steps (c) and (g), we can infer that the plots would be the same, if absolute wouldn't have been taken.
- Because of the absolute function, the part of the signal below the x-axis have come upwards.
- Both the plots would be the same, because we have applied the following operations to the function:
  - $ifft(exp(log(fft(voiced_{function})$
- Here, we can see that exp and log function are inverse of each other, and $ifft$ and fft are the inverse of each other too, thereby cancelling each other's effects and plotting the original voiced signal frame that we made earlier.

$Q2.$

(a) *Create a time domain plot.*



(b) *In the time-domain plot, mark the regions where the pitch is the highest and the lowest. What are the pitch frequencies in those regions?*



- Maximum pitch frequency = 0.5845 Hz
- Minimum pitch frequency = -0.2372 Hz
- Minimum pitch frequency if taken abs = 0 Hz

(c) *Write a code to calculate number of zero-crossing present in the signal.*

- Zero-crossing count came as 311.

(d) *Compute frame energy and comment on it.*

- [Done in MATLAB]
- Check the energy values from the workspace.
- These are the energy values

10x1 double

	1	2
1	0.0481	
2	0.1212	
3	4.5364	
4	10.9100	
5	16.5151	
6	13.9239	
7	5.6542	
8	17.0020	
9	3.1807	
10	0.2429	
11		

- Here, we are observing that the noise, unvoiced and silent part of the signal have energy values near zero, and the energy values much higher than 1 have high amplitude or frequency and thus represents the voiced part of the signal.

(e) *Comment on what type of region it is voiced or unvoiced.*

- If we observe the plot, the region where the amplitude of the signal with the frames of energy being high and number of zero crossings low, then that part of the signal or frame is voiced.

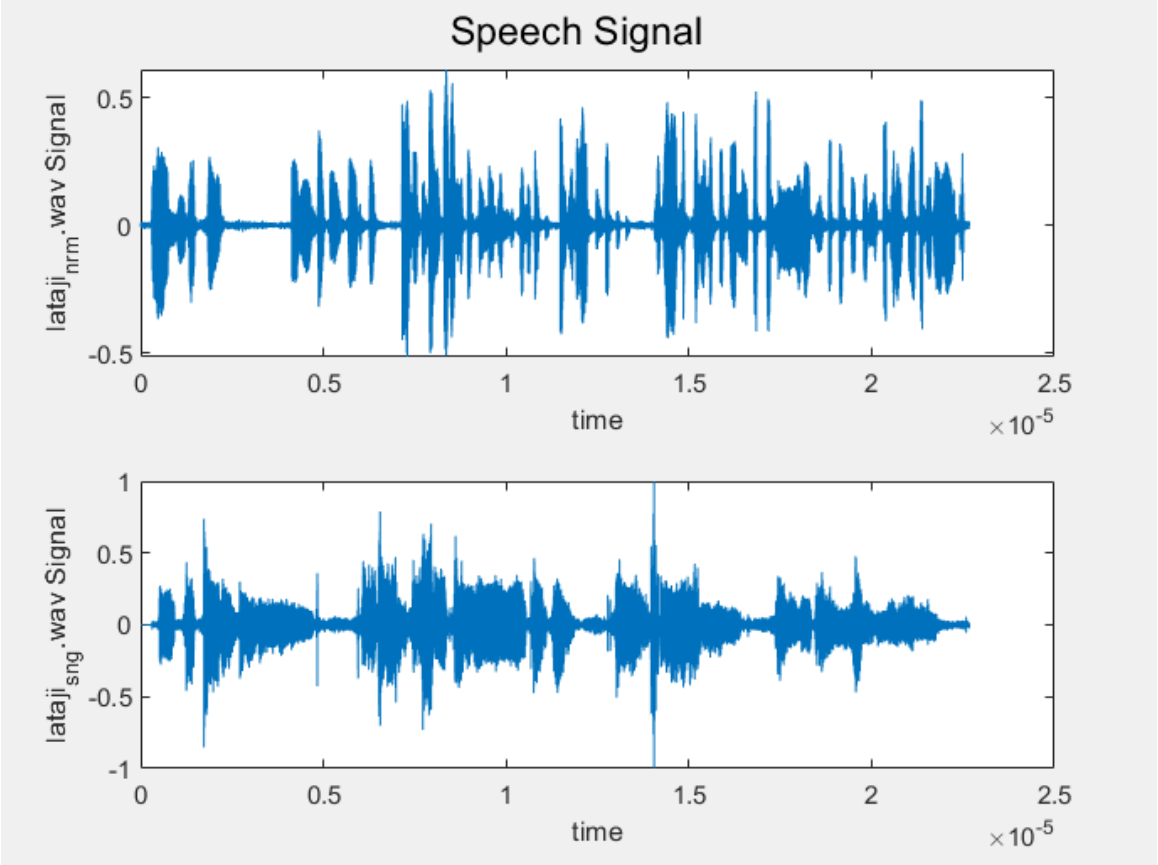
- If the amplitude of the signal and the energy of the frame is low, with high zero crossings, then, the region is unvoiced.

$Q3.$ Short note on the following: -

- **Zero-Crossing:** It is the point at which a function of a signal or otherwise crosses the horizontal axis as its value passes through zero and changes the sign.
- **Auto-correlation:** It is a mathematical representation that defines the correlation between the elements of a series and others from the same series separated from them by a given interval.
- **Epochs:** It is the instant of significant excitation of the vocal tract system during production of speech.
- **What are the assumptions to be considered while analyzing speech signal?**
    - We assume the signal is following linear time-invariant (LTI) system.
    - We also assume the signal to be stationary for a very short period of time, even though speech signals change very rapidly.

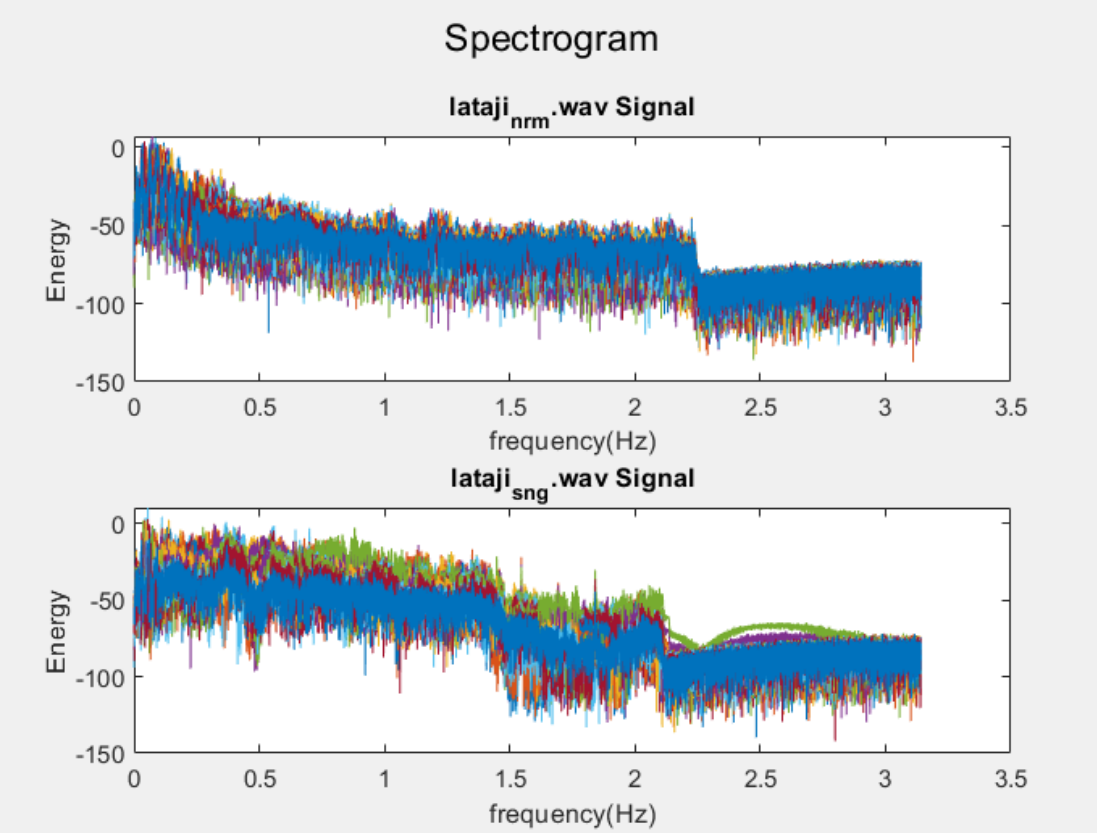
$Q4.$

(a) *Create time domain plots.*



(b) *Plot the pitch contour.*

- Spectrogram plots: Spectrogram represents the signal strength over time at various frequencies present in a particular waveform, and how energy levels vary over time.



(c) *Comment on the structure of pitch contour of both the signals.*

- The structure of pitch contour of both the signals show that the region of the signal with high energy, which is voiced part of the signal, is more concentrated or have higher amplitude in the pitch contour plot.
- The region of the signal where there is lower energy, which is silent and unvoiced part, will have lower amplitude in pitch contour plot (which is approximately zero).