EC5.408 Speech Signal Processing
IIIT Hyderabad
Monsoon 2022
2020102009

Assignment 2

Arshini Govindu

1. *Loads the audio file (H MKB.wav) which is provided to you into MATLAB/Python and perform the following task. The audio file is shared along with the assignment. Note: Computer-Based Question*

   (a) *Create a time domain plot (0.5 pts)*

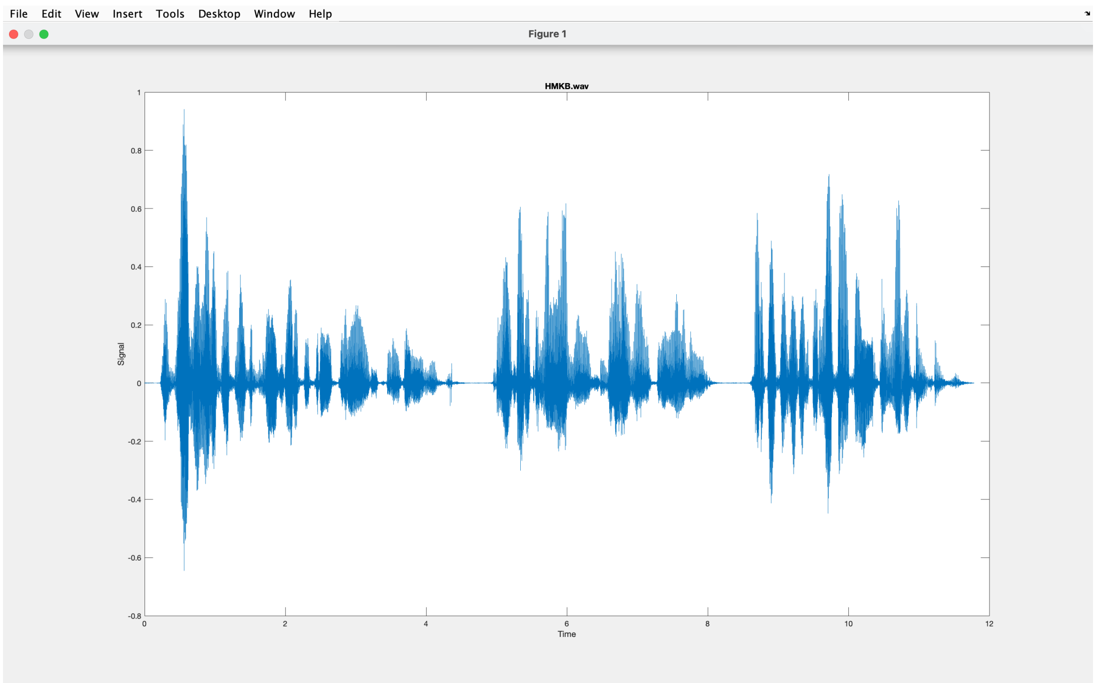

Figure 1: Time Domain Plot

   (b) *Apply framing onto the signal (0.5 pts)*

   (c) *Find one voiced frame among all the frames and plot it (0.5 pts)*

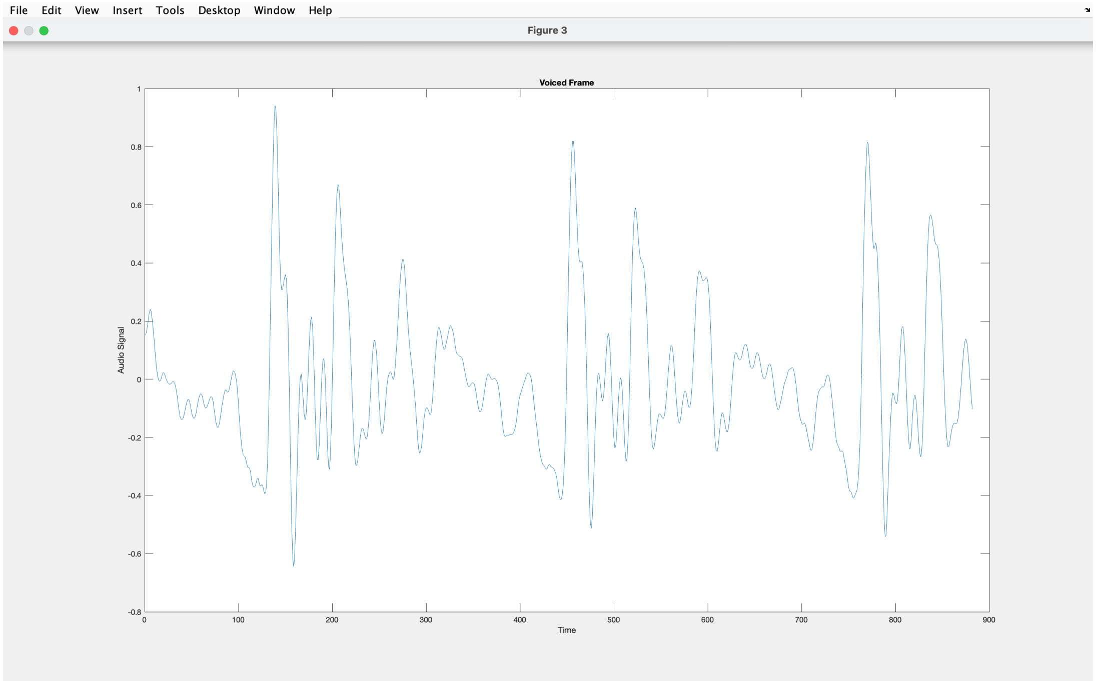

Figure 2: Voiced Frame Plot

(d) *Compute Fourier transform on the voice frame which you have considered in step (c) (0.5 pts)*

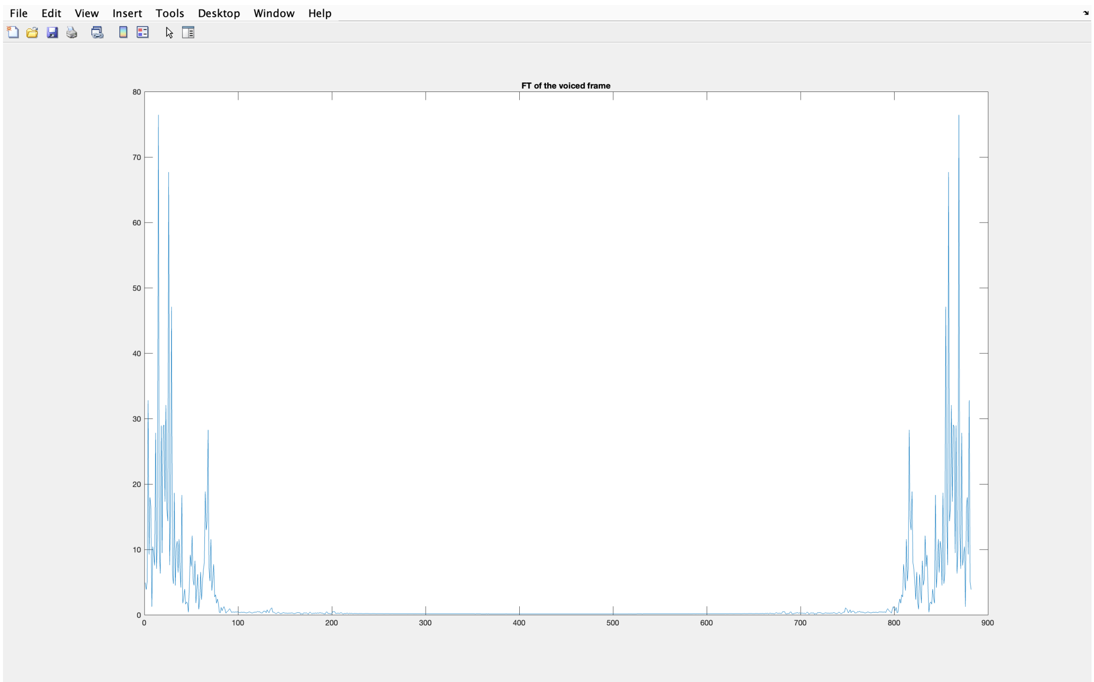

Figure 3: Time Domain Plot

(e) *Apply log to the step (d) and plot it (0.5 pts)*

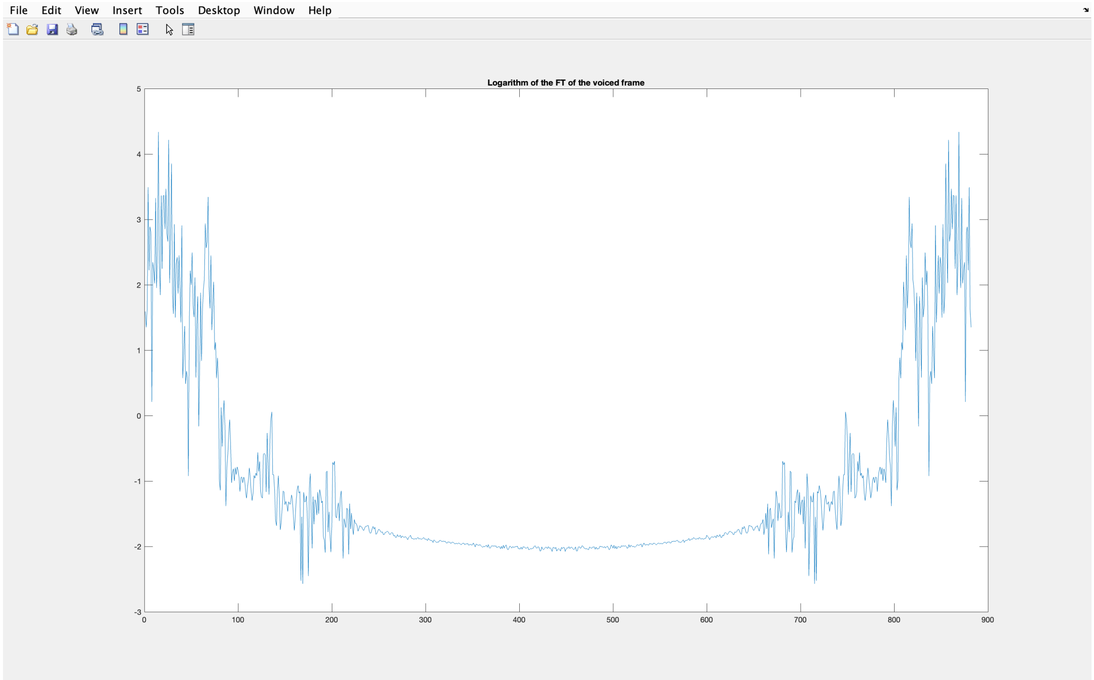

Figure 4: Time Domain Plot

(f) Compute exponential operation for the step (e)(0.5 pts)

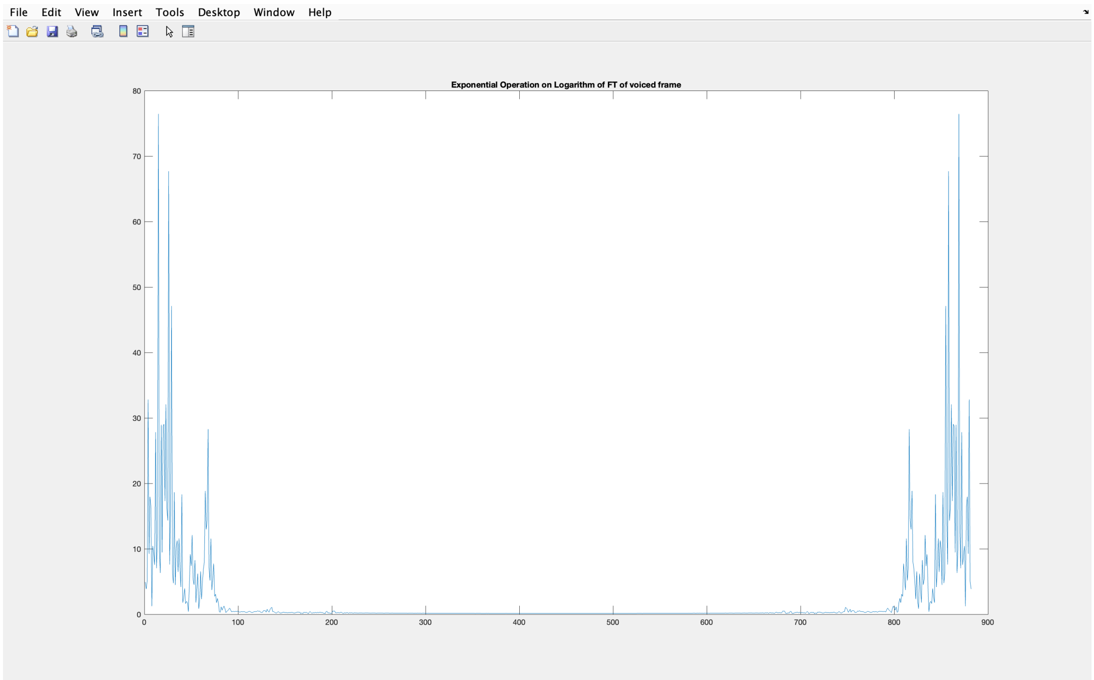

Figure 5: Time Domain Plot

(g) Compute Inverse Fourier transform for the step (f) and plot it (0.5 pts)

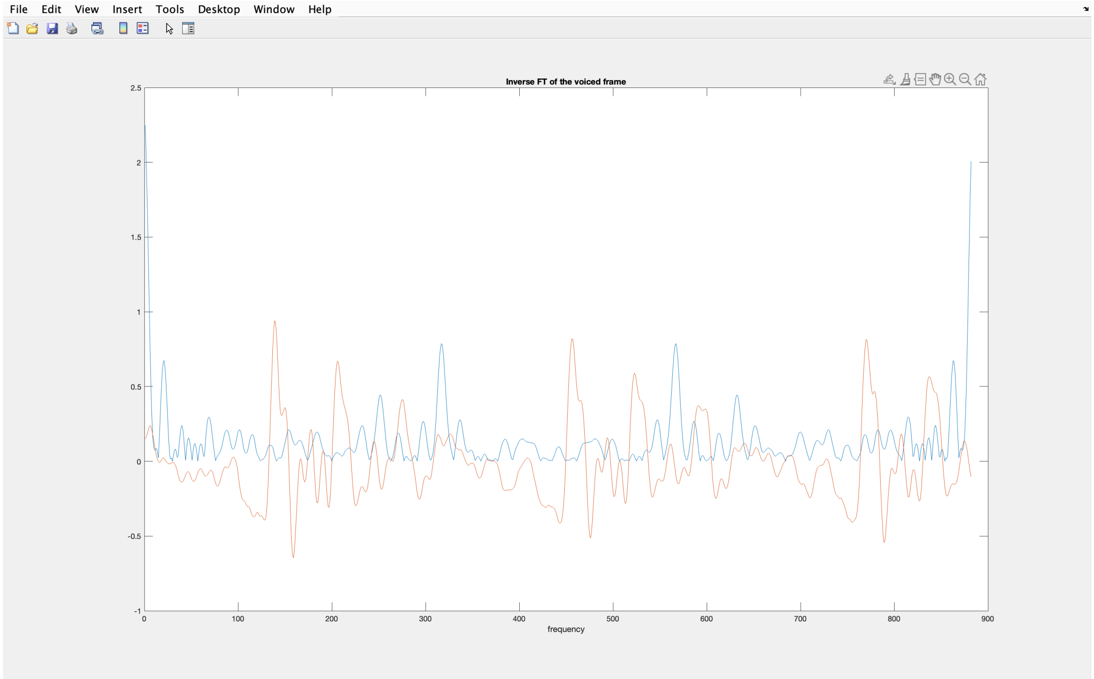

Figure 6: Time Domain Plot

The blue line represents the recovered signal after all the previous steps. The red line represents the original unmodified frame that we selected

(h) *Comment on step (c) and step (g). If the outputs are the same justify why. vice-versa (1.5 pts).*

The reason why it doesn't have the same outputs is because for the calculation of log and exponent after that we have to take the absolute values of the magnitude spectrum. The complex values are removed for ease of calculation.

2. *Load the audio file (chunk1.wav) which is provided to you into MATLAB/Python and perform the following task. The audio file is shared along with the assignment. Note: Computer-Based Question*

   (a) *Create a time domain plot (0.5 pts)*

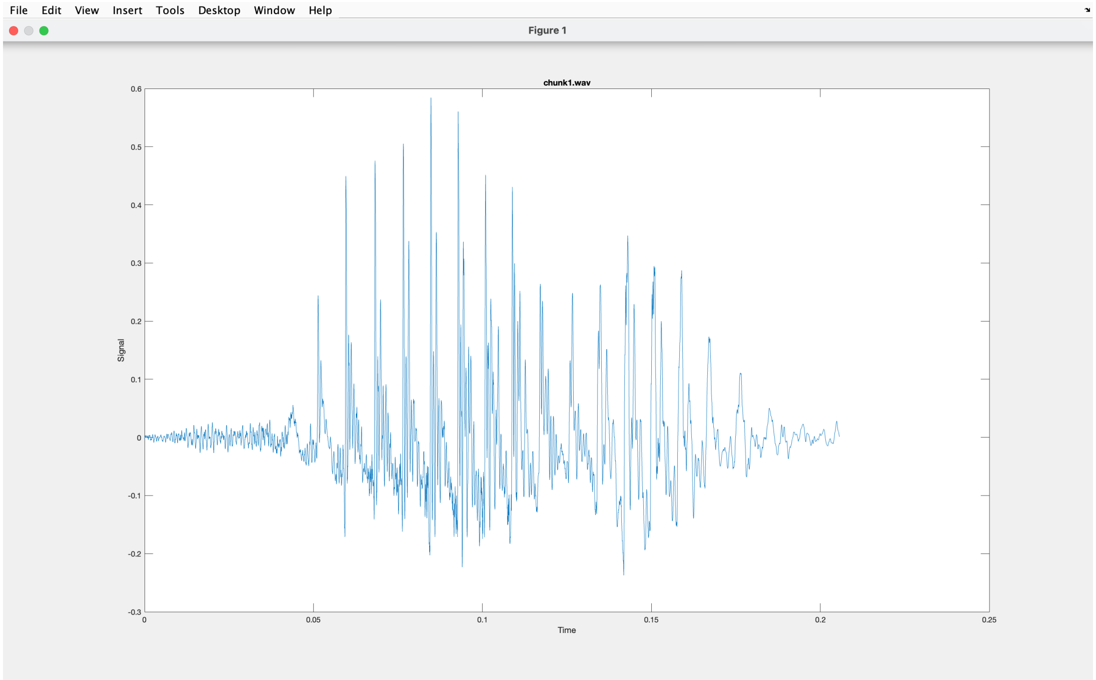

Figure 7: Time Domain Plot

(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? (0.5 pts)*

The pitch frequencies in those regions are:

i. Highest = 133.72Hz

ii. Lowest = 116.05Hz

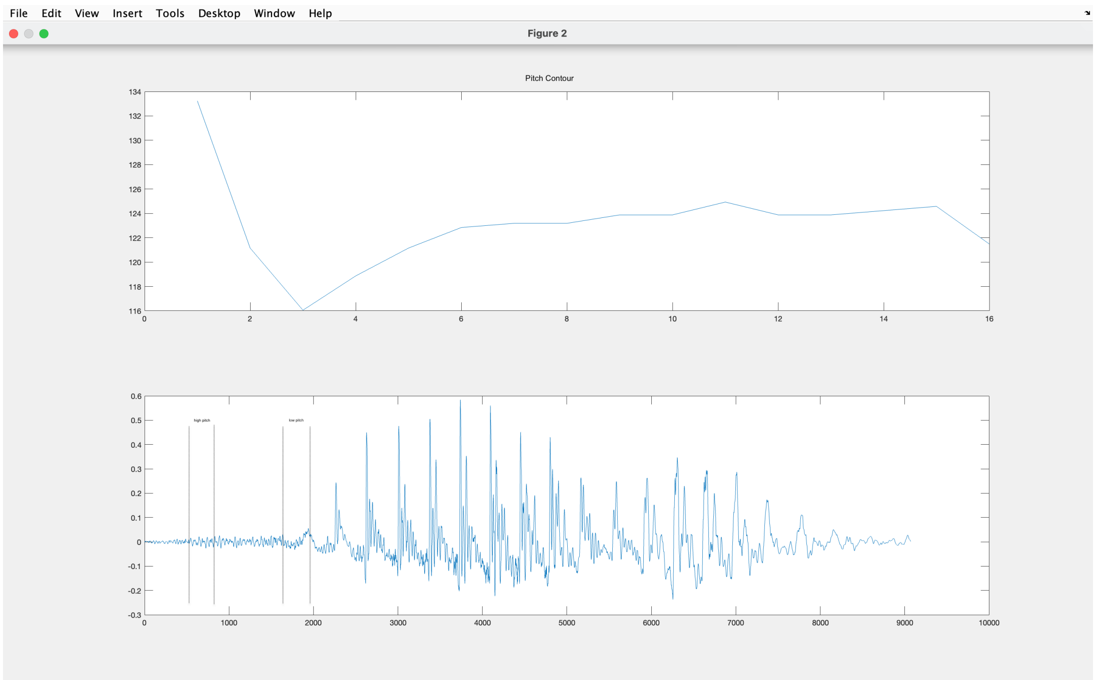

Figure 8: Time Domain Plot

(c) *Write a code to calculate the number of zero-crossing present in the signal (2 pts)*

```
ZC = 0;
for k = 1:length(y)-1
if ((y(k) < 0) && (y(k + 1) > 0 ))
ZC = ZC + 1;
elseif ((y(k) > 0) && (y(k + 1) < 0))
ZC = ZC + 1;
end
end
```

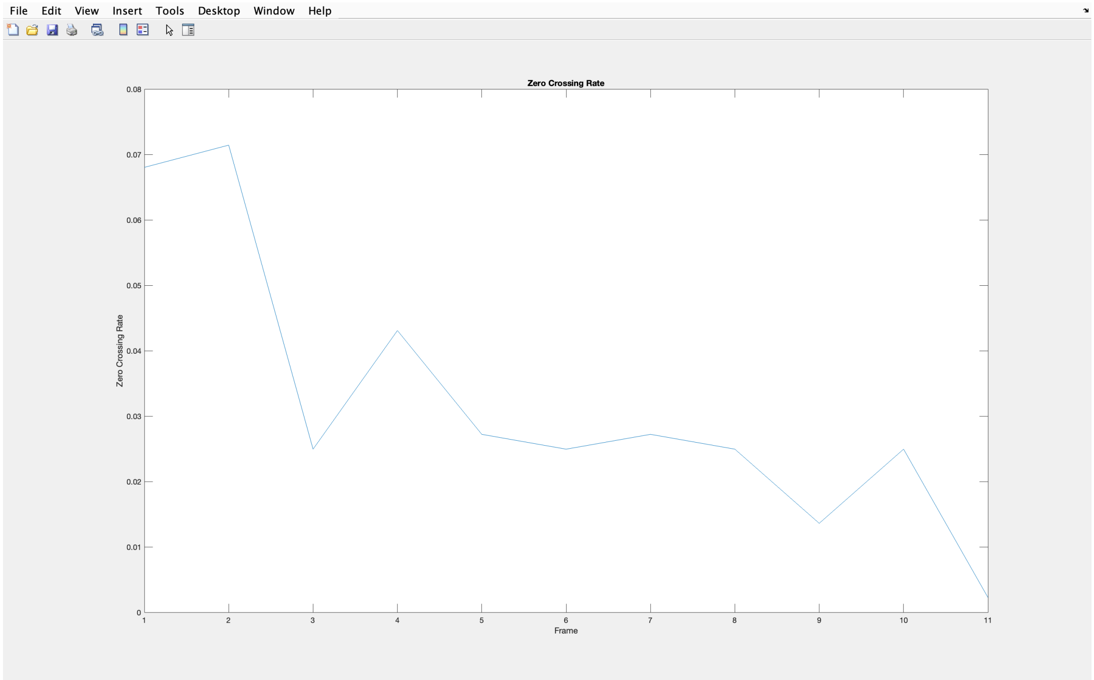

Figure 9: Time Domain Plot

(d) *Compute frame energy and comment on it (1 pts)*

We can calculate frame energy as

$$E_x = \sum_{n=-\infty}^{\infty} |y[n]|^2 \tag{1}$$

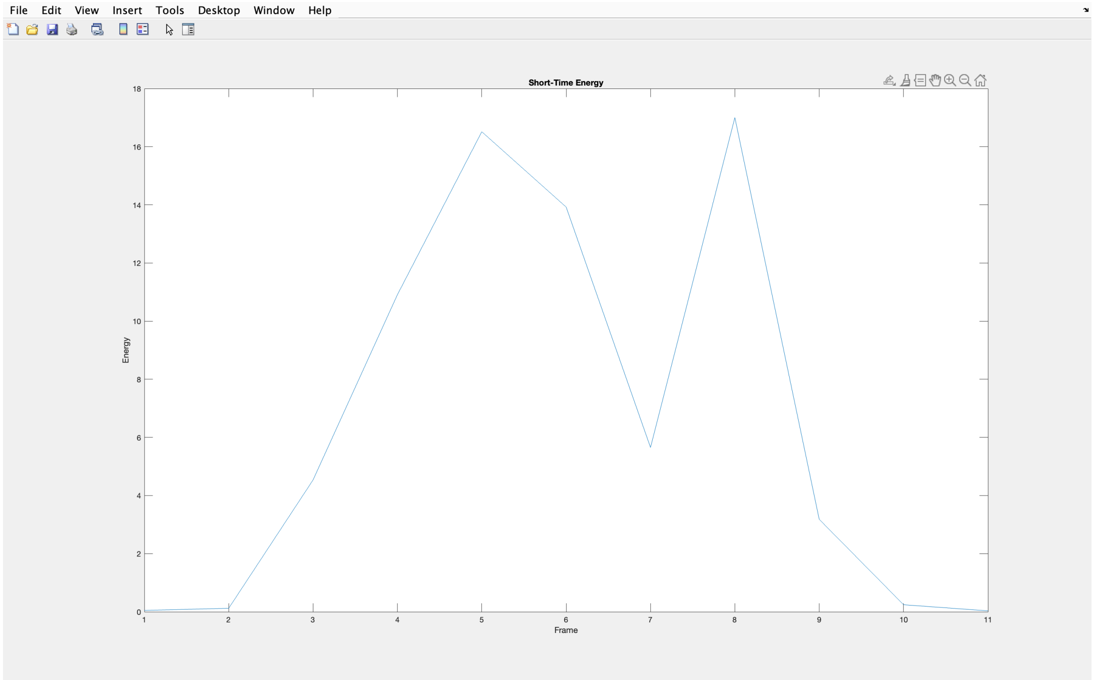

Figure 10: Time Domain Plot

(e) *Comment on what type of region it is voiced or unvoiced (1 pts)*

The parts of the graph where the frame energy is more are the regions where the speech is voiced.

We know that the frame energy is directly proportional to the square of the amplitudes of the signal. So we can say that voiced speech has more energy.

So the parts with lower energy are unvoiced speech. We can decide whether or not a frame is voiced speech or unvoiced by setting a threshold on the energy. Frames with lower energy than the threshold value is considered to be unvoiced and those above it are said to be voiced parts of speech.

Analyzing the signal using the zero crossing rate, in unvoiced speech, there is a lot of randomness so that part would have more ZCR than voiced speech which has some sort of periodicity at least. The voiced speech regions have lower ZCR compared to unvoiced speech.

3. *Write a short note on the following:*

(a) *Zero-Crossing (1 pts)*

The Zero-Crossing Rate of a frame is the rate of change of signs of the signal in the frame. It is the number of times the signal changes value, from positive to negative and vice versa, divided by the length of the frame. It usually exhibits higher values in the case of signals with noise. Voiced speech would have lesser ZCR than unvoiced speech and noisy speech.

(b) *Autocorrelation (1 pts)*

Autocorrelation is the correlation of a signal with a time-delayed version of itself. If a signal is periodic then the signal will be perfectly correlated. This leads us to understand that the autocorrelation values help to figure out whether a signal buried in noise is periodic or not.

$$R(\tau) = \frac{1}{T}\int_0^T s(t)s(t+\tau)dt \tag{2}$$

(c) *Epochs (1 pts)*

The instant of significant excitation of the vocal-tract system during production of speech is called epoch. The most significant excitation takes place around the instant of glottal closure. An excitation is represented by an impulse as these are usually much more than the surrounding impulses. It is also called as glottal closure instant. Epoch locations are useful for determining the fundamental frequencies of the speech signal.

(d) *What are the assumptions to be considered while analyzing speech signals? (2 pts)*

While analyzing speech signals, we make a considerable number of assumptions as a speech signal is not completely stationary and periodic.

i. We assume that a speech signal is quasi-stationary i.e. small segments of a speech signal called frames that are of 20ms duration approximately are considered to be stationary.

ii. Speech production is excitation source that is passed through a linear digital filter. The excitation source represents either voiced or unvoiced speech, and the filter models the effect produced by the vocal tract on the signal.

iii. Speech production can be modeled as an LTI system if we take a short time segment of the speech signal.

iv. The excitation source is the air from the lungs that is released suddenly after it is stored up so it can be assumed to be an impulse.

v. The vocal tract system is considered to be an all-pole system with the poles representing the formants if we consider a linear prediction filter.

Your answer goes here

4. *Load the audio files lataji nrm.wav and lataji sng.wav into MATLAB/Python and per- form the following task. The audio file is shared along with the assignment. Note: Computer-Based Question*

(a) *Create a time domain plot (1 pts)*

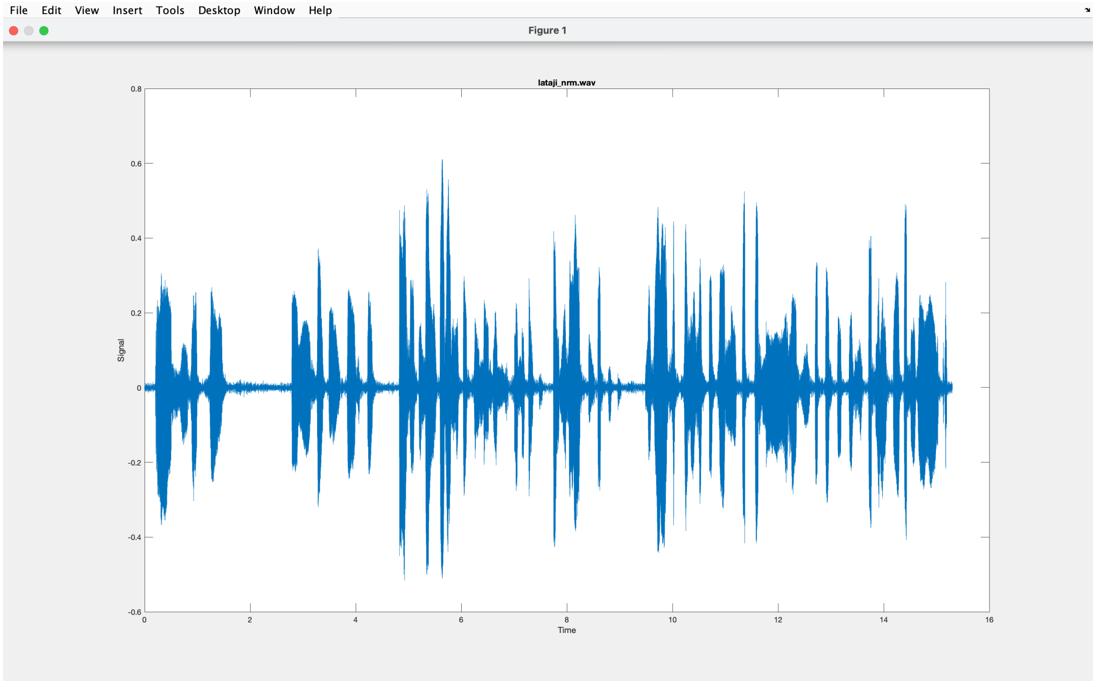

Figure 11: Time Domain Plot of lataji-nrm.wav

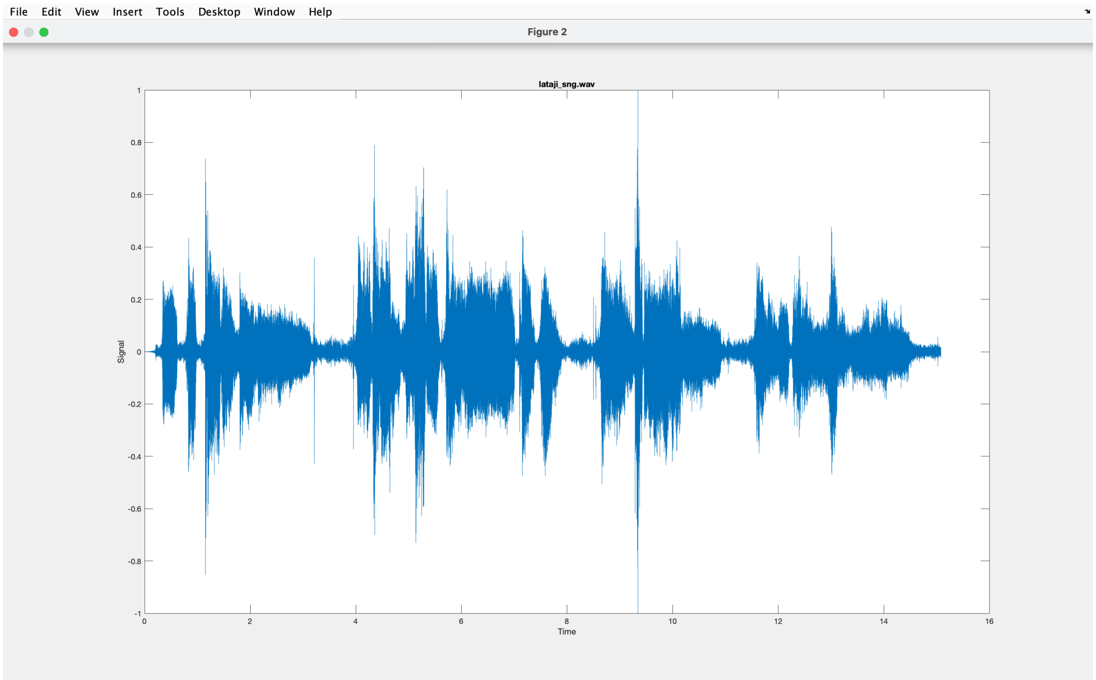

Figure 12: Time Domain Plot of lataji-sng.wav

(b) Plot the pitch contour (2 pts)

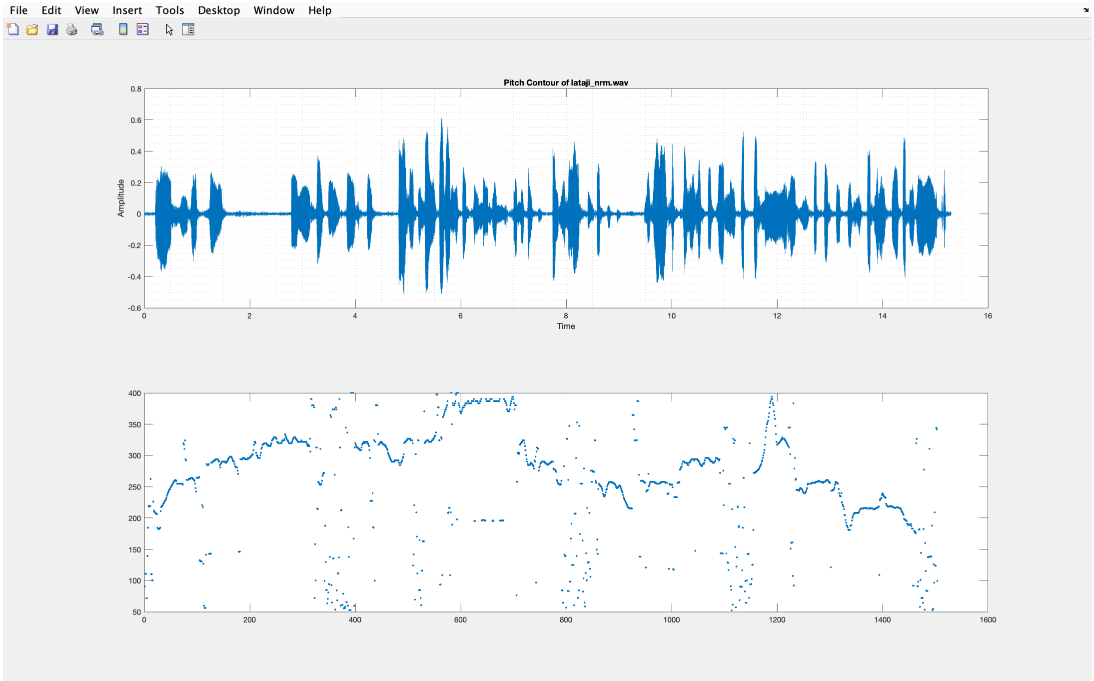

Figure 13: Pitch Contour of lataji-nrm.wav

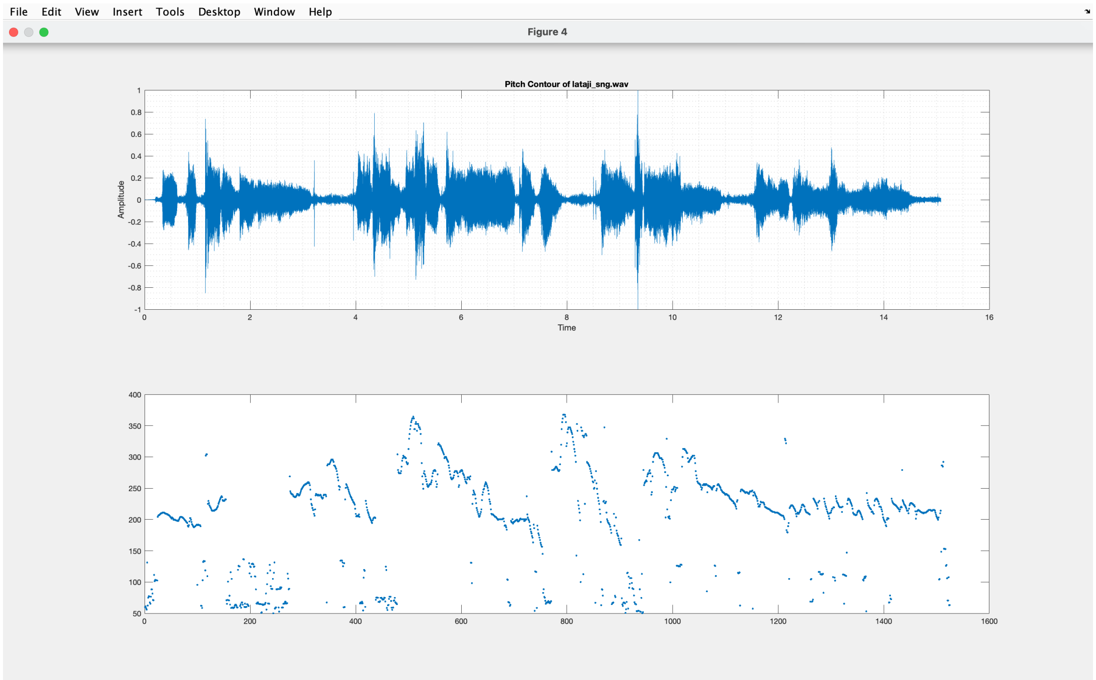

Figure 14: Pitch Contour of lataji-sng.wav

(c) *Comment on the structure of pitch contour of both signals. (2 pts)*

lataji-sng.wav has more number of voiced regions than the lataji-nrm.wav as we can see from the pitch contour. The lataji-sng.wav file has more periodic pitch and this means that there are more regions with periodicity. So that is why it has more pitch.