EE679 Assignment 2

Name : Agulla Surya Bharath   **Roll.No : 17D070055**
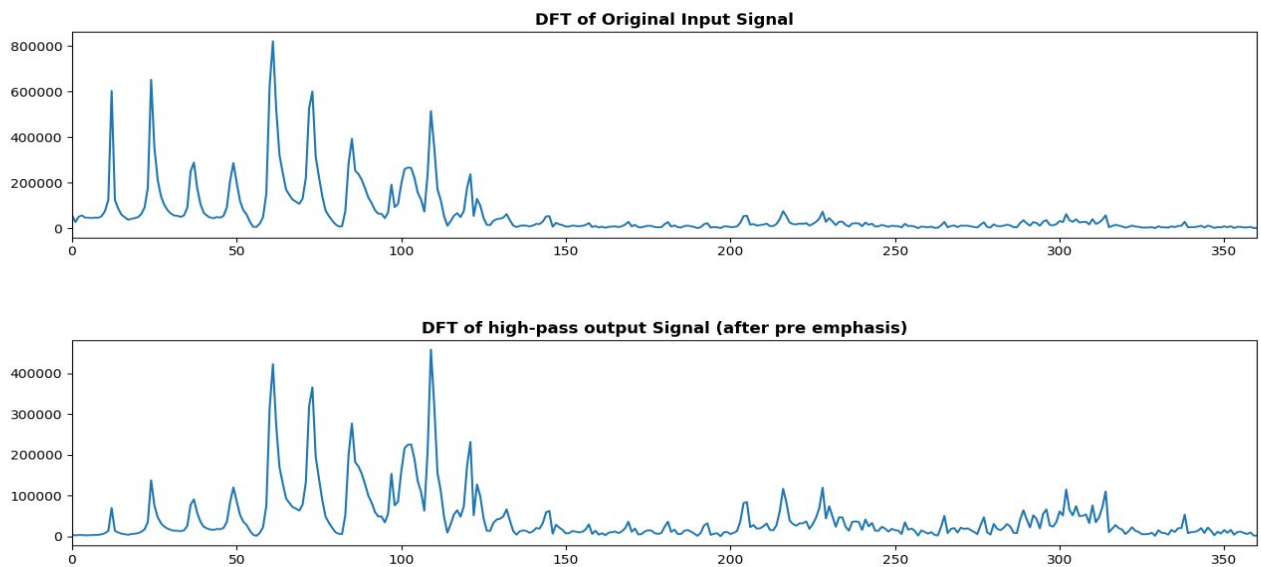
Code file for whole assignment : "assmt2.py"

Q1) Pre emphasis of Input Signal
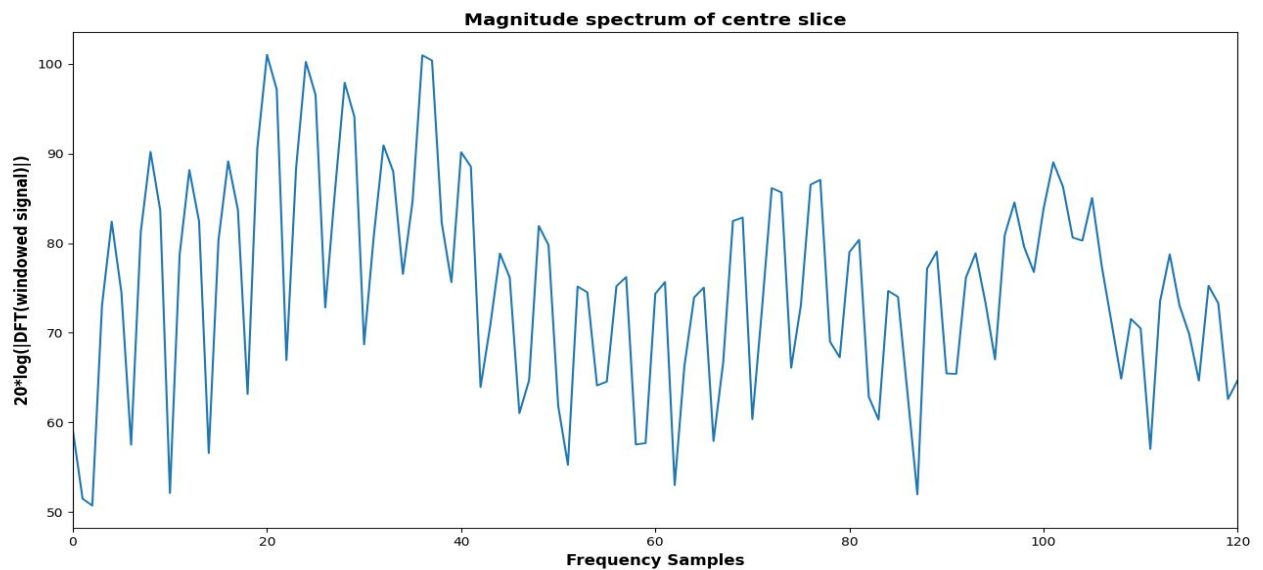
Filter(z) = 1- αz^{-1}

I took $\alpha = 0.95$

Pre emphasis of Input

DFT of Original Input Signal

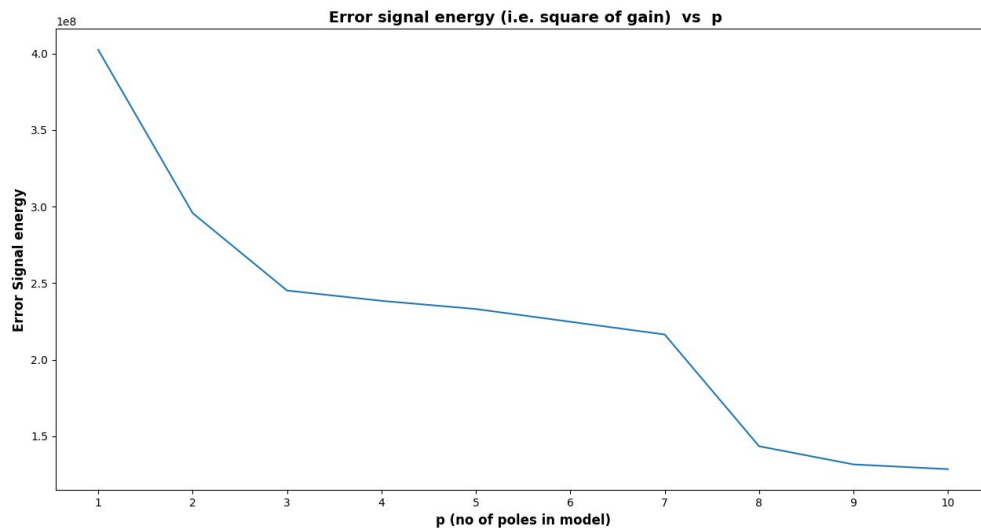
DFT of high-pass output Signal (after pre emphasis)

Audio Signal : " pre_emp.wav "

Q2) Narrowband magnitude spectrum slice using a Hamming window of 30 ms on a segment near the centre of the given audio file
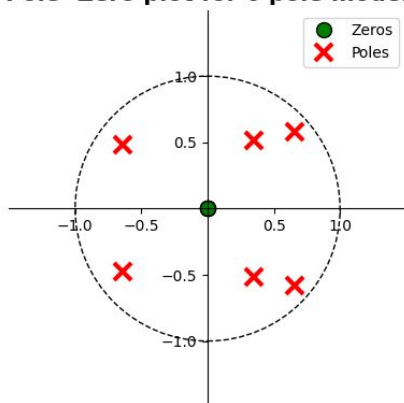
Magnitude spectrum of centre slice

Q3)Plot error signal energy (i.e. square of gain) vs poles (p)

The Error signal energy is decreasing as the number of poles in the model increases. This is inline with our expectation, as discussed in class

Q4) Pole zero plots for p = 6 and p = 10

It is evident from above pole zero plots that all poles of the model transfer function lie inside the unit circle and hence we can conclude that our model is stable.

Q5) LPC spectrum magnitude (i.e. the dB magnitude frequency response of the estimated all-pole filter) for each order "p".

The following the actual unshifted version of the plot

**dB magnitude frequency response of the estimated all-pole filter**
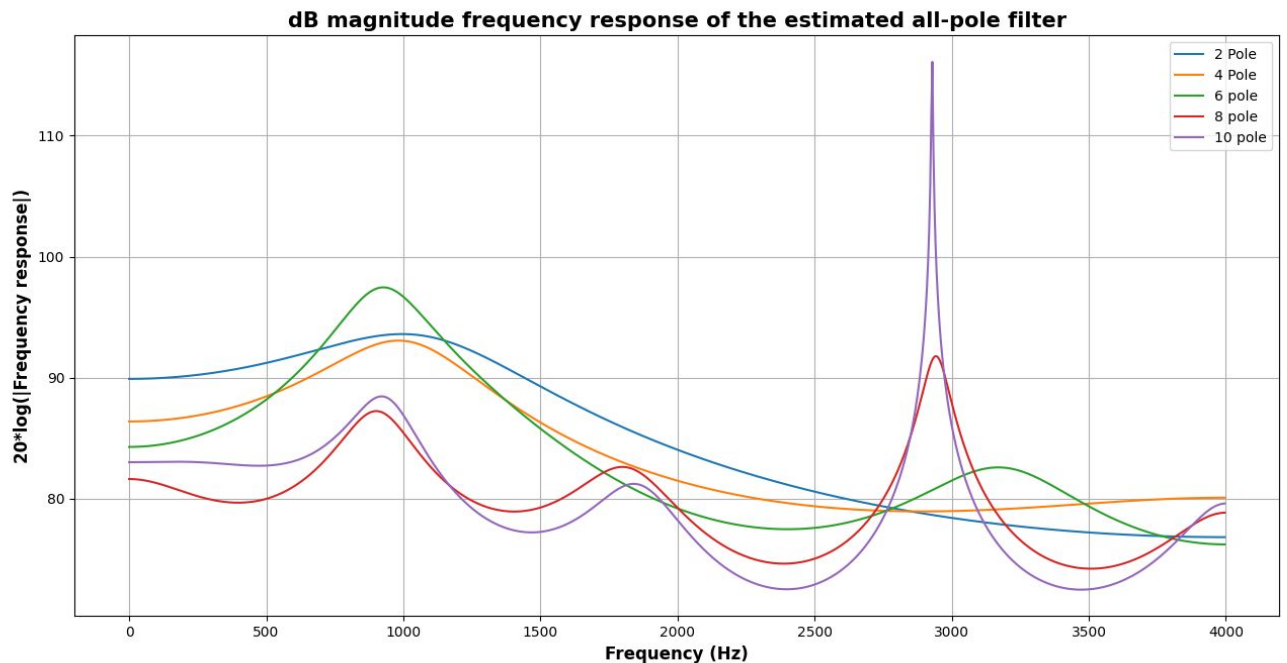
        Inorder to get the view of comparison between the filters of different poles, i have shifter each plot above vertically and achieved the following plot


dB magnitude frequency response of the estimated all-pole filter
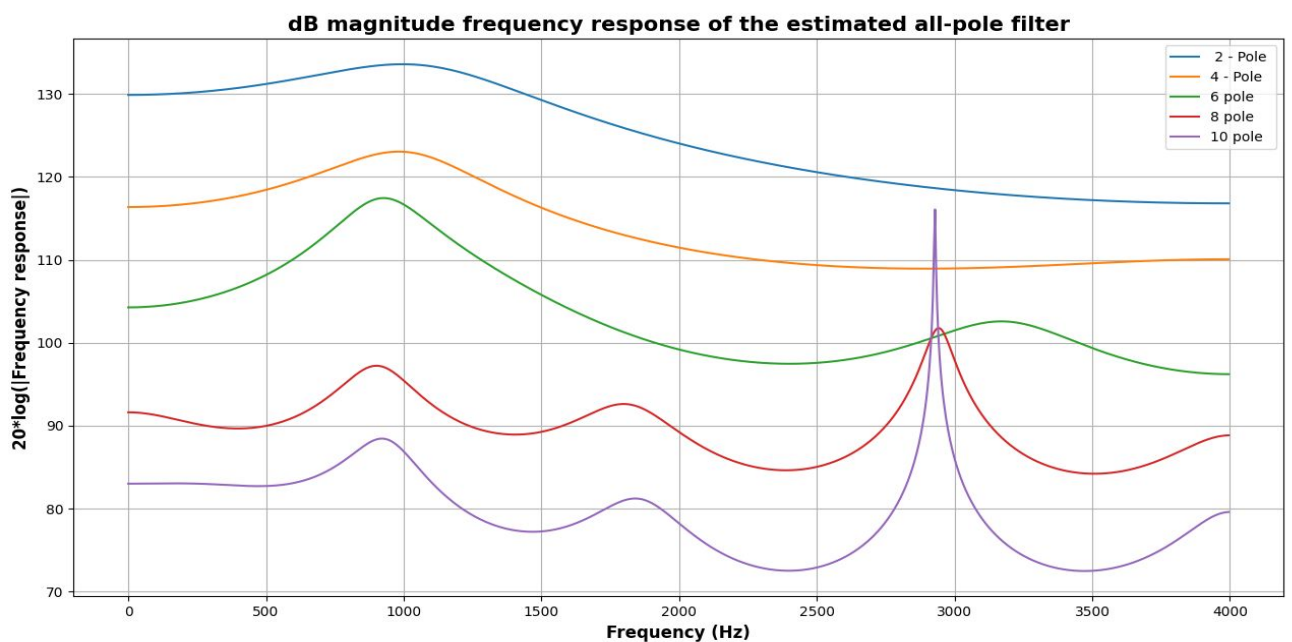
Comment on the characteristics of the spectral envelope estimates.

Our theoretical expectation is that for a 2n-pole model we must have n - peaks in the frequency response plot . Here the plots do seem to have quite similar behaviour , also as the number of poles in the model increases we can extract more information about the true vocal tract filter characteristics like peak positions and relative amplitude gains at peaks and also width / sharpness of different peaks.

**Comment on their shapes with reference to the short-time magnitude spectrum computed in part 2.**

Magnitude spectrum of centre slice
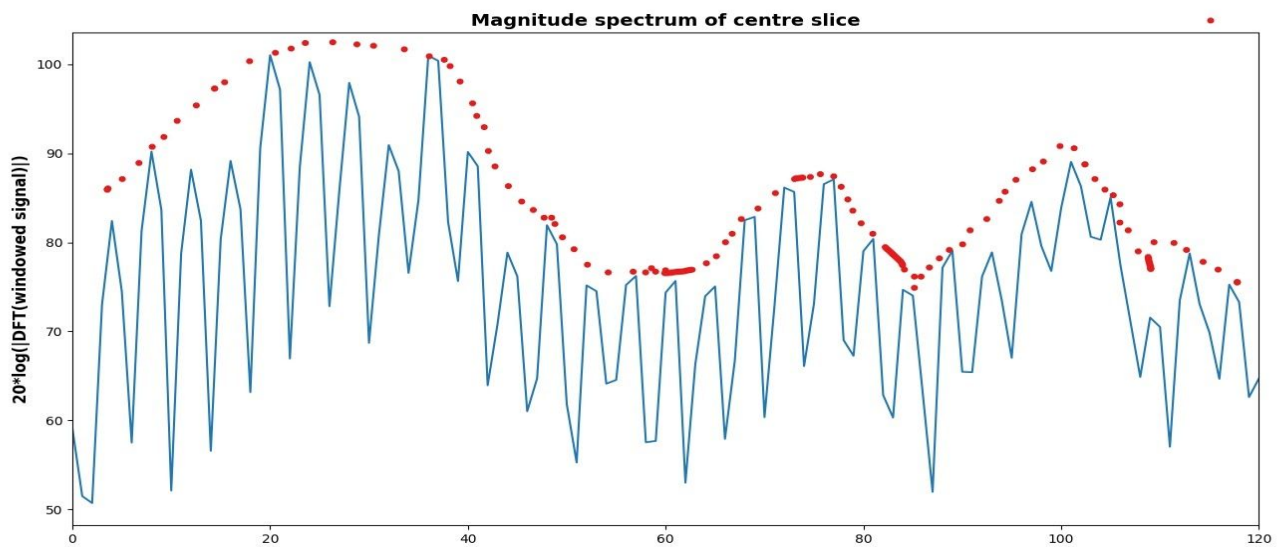
This above figure shown rough envelope structure

dB magnitude frequency response of the estimated all-pole filter
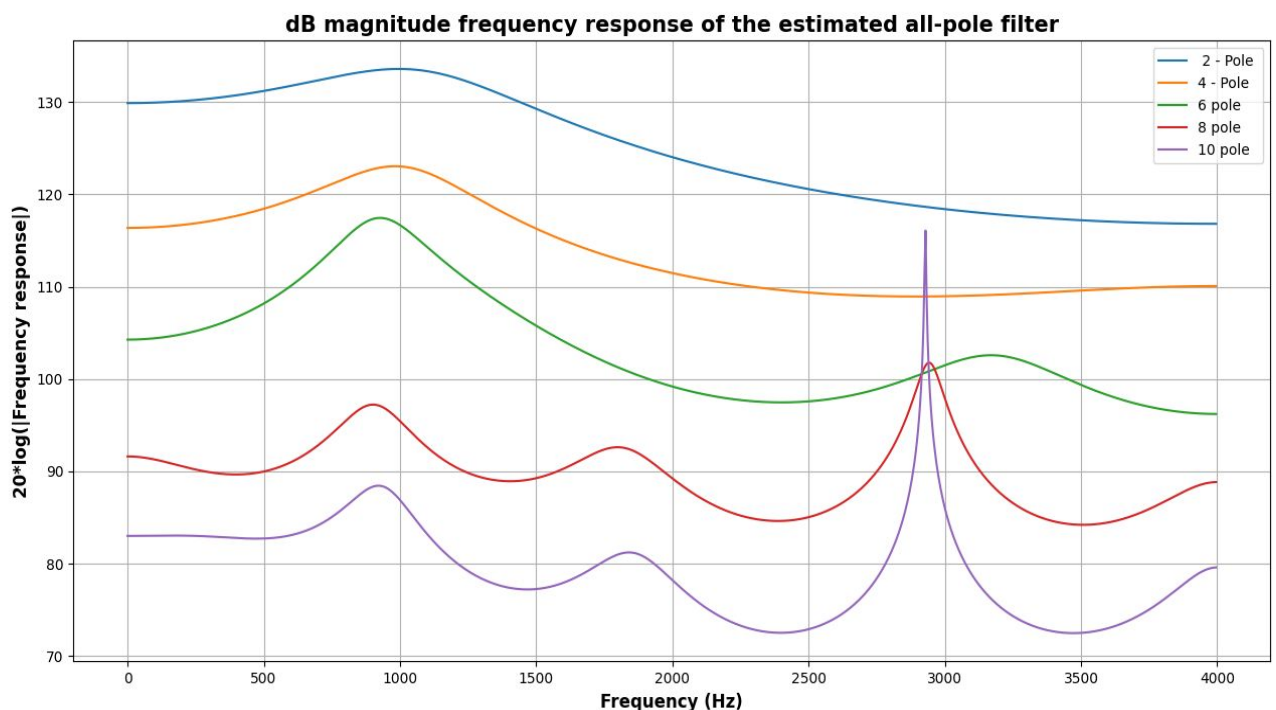
Now comparing the number of peaks and relative positions and relative amplitudes of peaks in the envelope with the peaks in frequency response plots of all pole filters will leave us a strong clue that the actual vocal tract might be a very close to 8-pole or 10-pole model created by us using LP analysis. So, the vocal tract cannot be a 2, 4, 6 pole filter .

Q6) Measuring pitch period

Residual Error signal from Inverse filtering

Autocorrelation of Residual Error signal
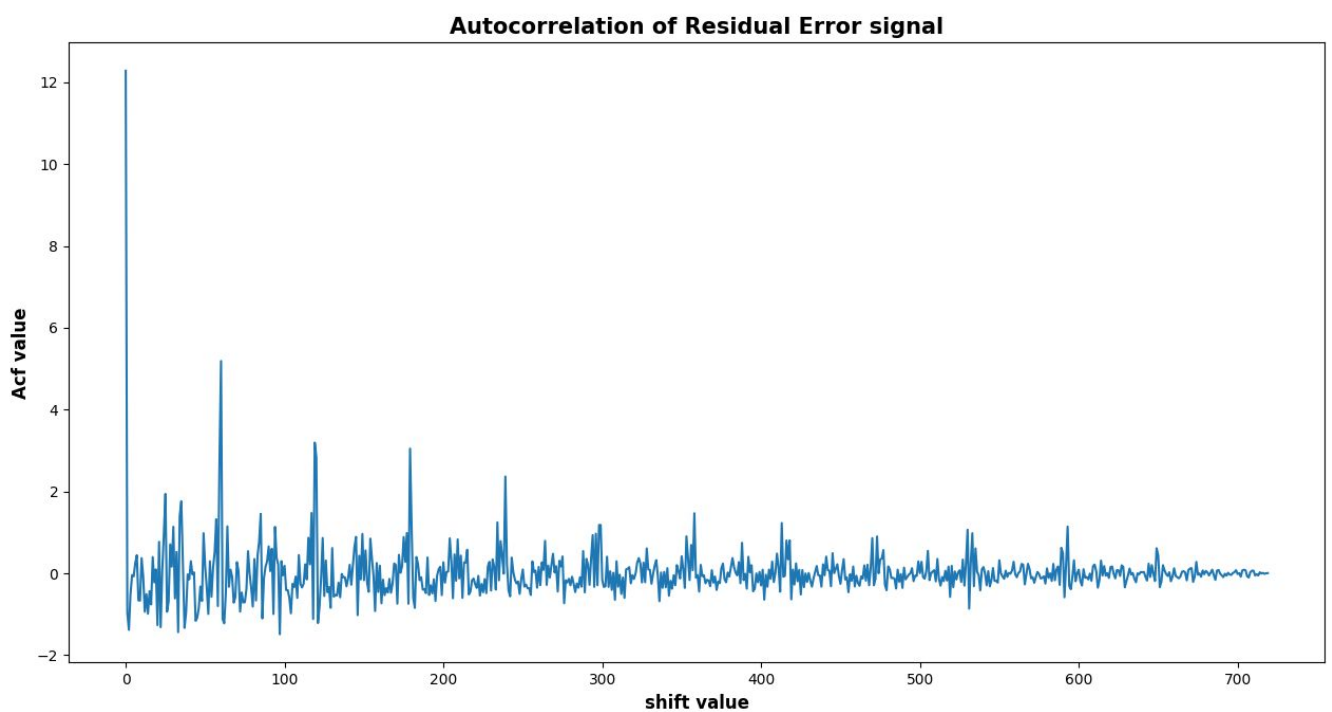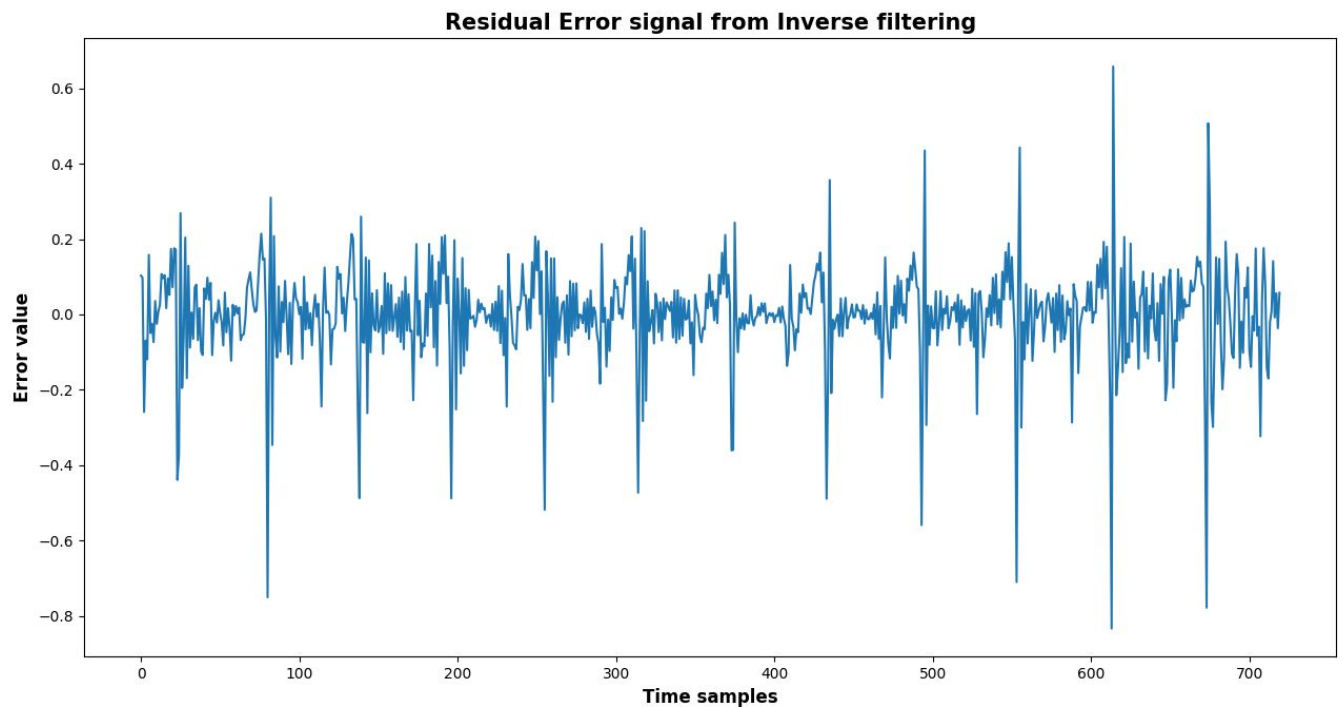
$$\text{Acf[shift value]} = \Sigma_n \text{ error[n] * error[n+shift value]}$$
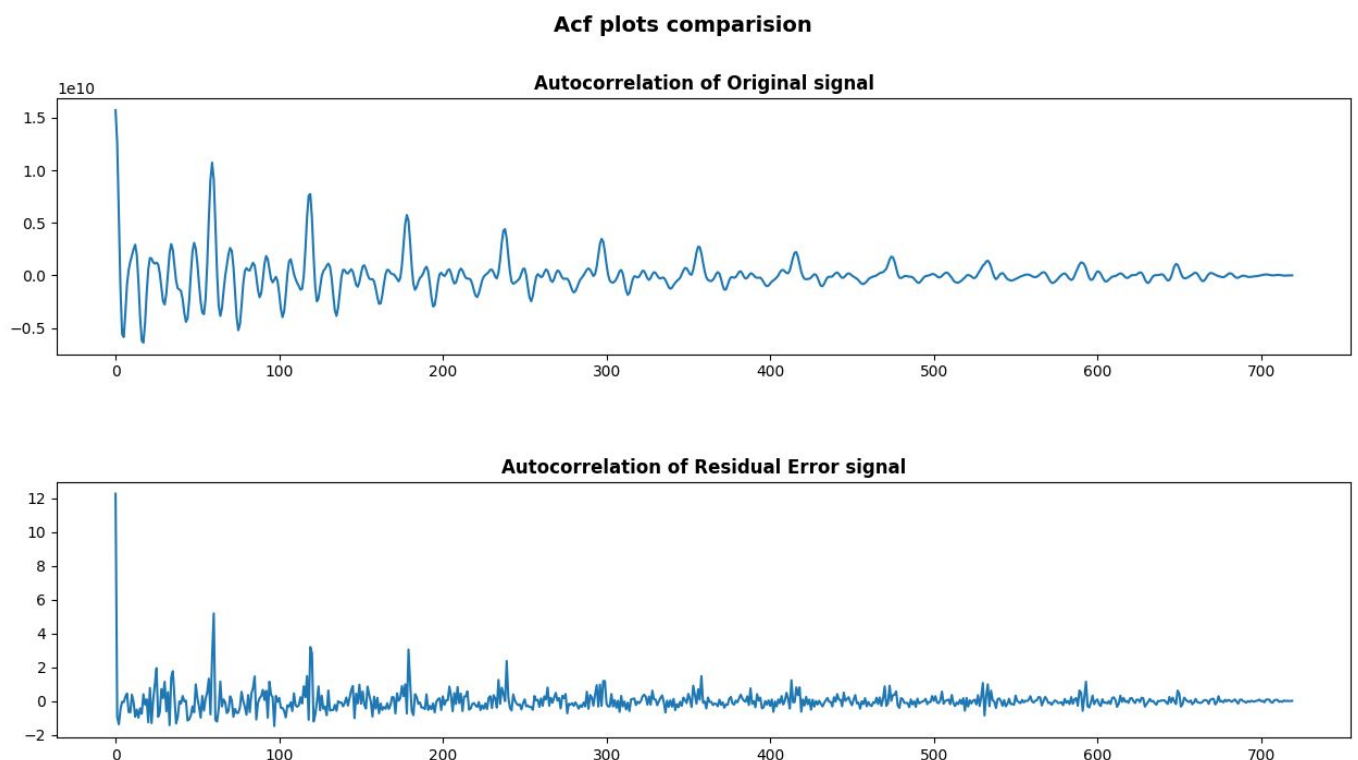
The peaks in acf plot of Residual error signal are at 0 , 60 , 120 , 180 ,.....
Shift values in calculating the acf .

Thus the pitch period of the original input signal is 60 samples.
We know sampling frequency = 8000 samples/sec

Thus pitch period = 7.5 ms

Hence fundamental frequency = 133.3 Hz

Comparing the Acf of original signal and residual error signal

Acf plots comparision

Autocorrelation of Original signal

Autocorrelation of Residual Error signal

Autocorrelation of Original Signal and Residual error signal are closely looking similar in the position of peaks at pitch periods, this is expected.

Also we can clearly say that the acf of residual error signal is more convenient and reliable plot to get the pitch period as compared to acf of original signal because acf of original signal has more peaks surrounding the pitch period peak(may due to formant structure) and where as in acf of residual error these extra peaks got suppresses significantly.
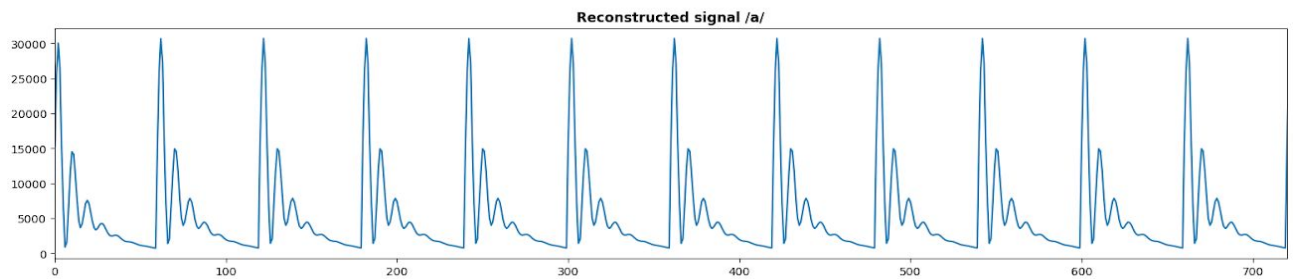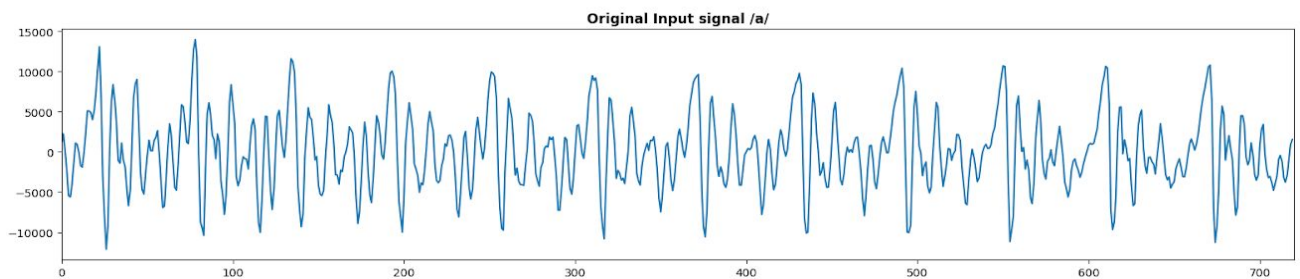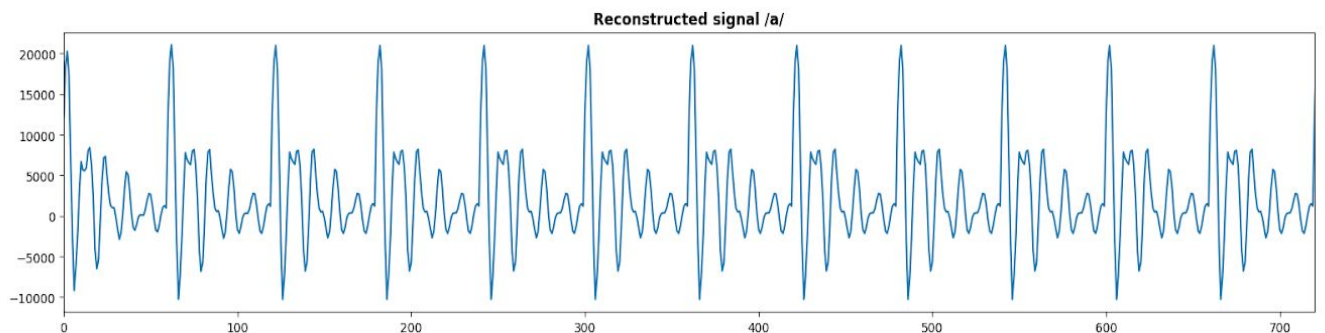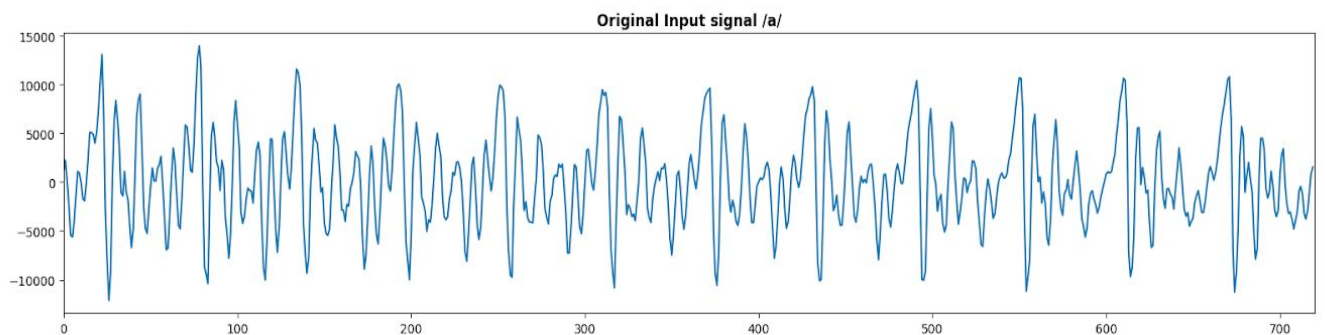
Q7) LP re-synthesis

For the final step in re-synthesis (i.e de-emphasis) part I have used
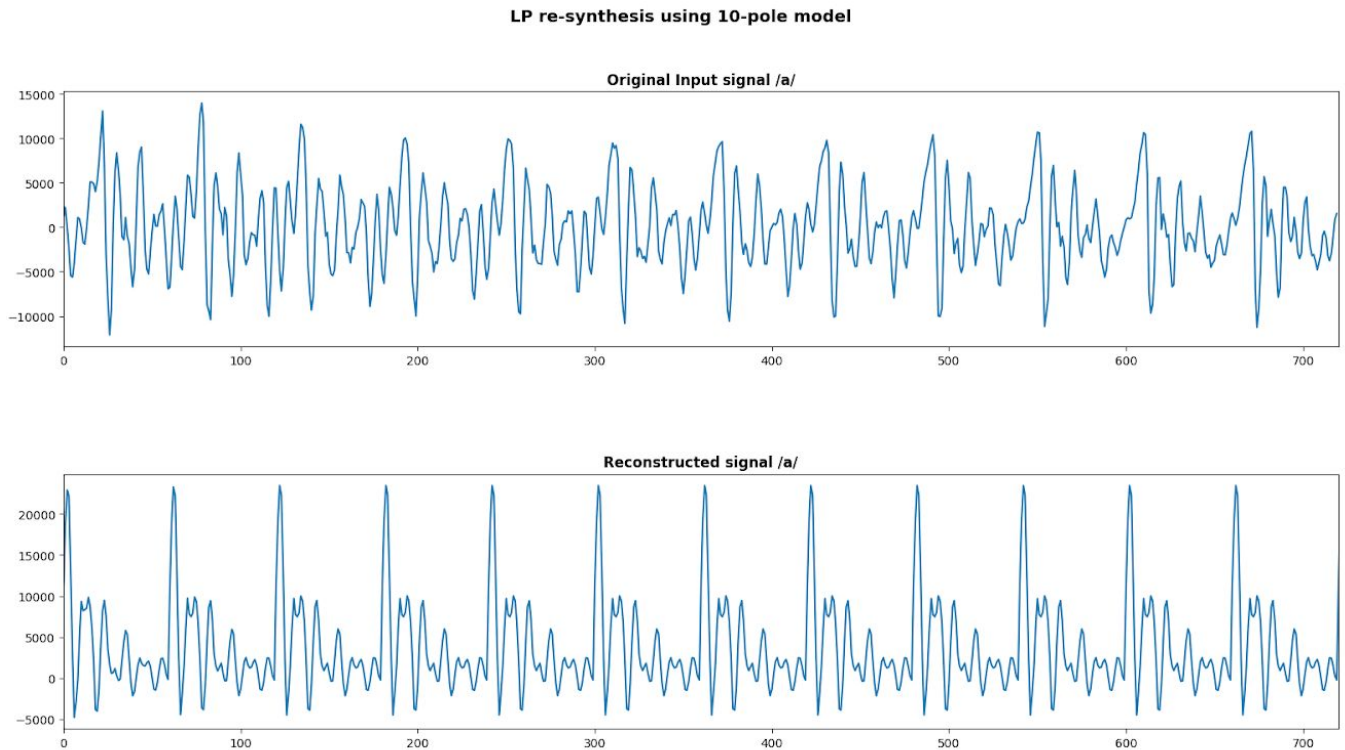Filter(z) = $1/(1- \alpha z^{-1})$ and took α = 0.95

For the re-synthesis part I have generated multiple outputs for different pole
models and have shown the corresponding plots below.

LP re-synthesis using 6-pole model

Original Input signal /a/

Reconstructed signal /a/

LP re-synthesis using 8-pole model

Original Input signal /a/

Reconstructed signal /a/

Original Input signal /a/

Reconstructed signal /a/

Both **8-pole model** and **10-pole model** seem to give a closely better approximation of original vocal tract ..

Between those two **10-pole model** gives more better approximation , This is evident by looking at the reconstructed signal and original signals in the above plots…

Note : Given input file "aa.wav" is only of 90ms duration but I have synthesized the /a/ sound for 300ms as asked in question and so I have assumed that the parameters like LP_coefficients of model , G to be remaining same for this 300ms.

But in reality we do this analysis for small duration like 30ms where the sound signal has only one vowel in whole duration and hence we expect the statistical parameter to not change much in this small duration.

Still , we have achieved a very good approximation of the original signal and also we can clearly notice the vowel from the synthesized sound (using a 10-pole model).
So, this makes us witness the fact that speech sound signals can be encoded in less number of bits (only Lp coefficients and G value ,pitch period) than actual bits of quantizing the samples.

Audio files: For 6 - pole   : "synthesized_aa_6.wav"

For 8 - pole   : "synthesized_aa_8.wav"

For 10 - pole : "synthesized_aa_10.wav"

We can clearly perceive the vowel in "synthesized_aa_10.wav" to be /a/