Speech Processing
EE679

Mohit
20D070052

October 2023

1 Student Details

Name: Mohit
Roll No: 20D070052

2 Question 1

Apply pre-emphasis to the signal.
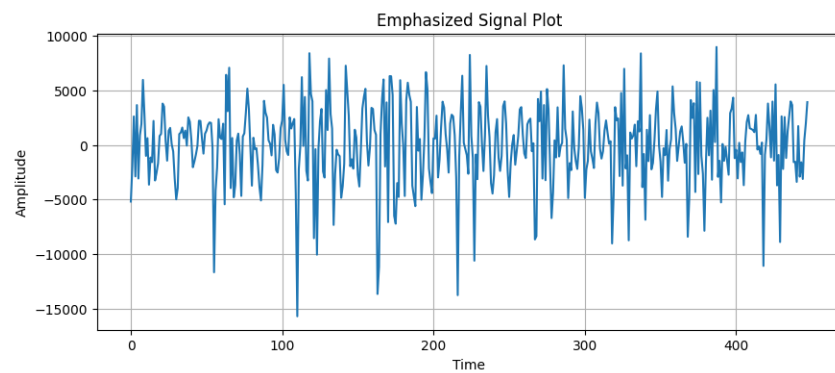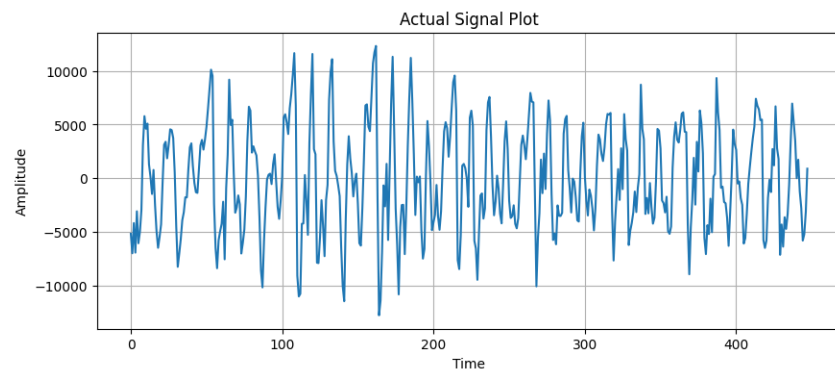
2.1 Solution 1

Pre-emphasis is used as a high-pass filter with the formula P(z) = 1- az-1, where 'a' is a parameter, to slightly uniformalize the spectrum of the voice input, here I have used 0.97 as default value

2.2 Code

The code files are included in the .ipynb file included in the submission.

2.3 Plots

3 Question 2

Next, compute and plot a single narrowband magnitude spectrum slice using a Hamming window of duration = 30 ms on a segment near the centre of the given audio file.
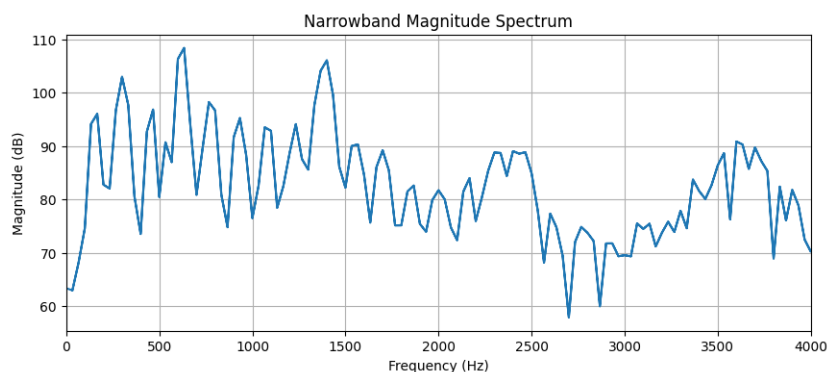
3.1 Solution 2

Directly using the np.hamming function present in the numpy library for this.

3.2 Code

The code files are included in the .ipynb file included in the submission.

3.3 Plots

4 Question 3

With the same 30 ms segment of part 2, compute the autocorrelation coefficients required for the LPC calculation up to p=10. Use the Levinson-Durbin recursion to compute the LP coefficients from the autocorrelation coefficients for each order p = 2,4,6,8,10. Plot error signal energy (i.e. square of gain) vs p.
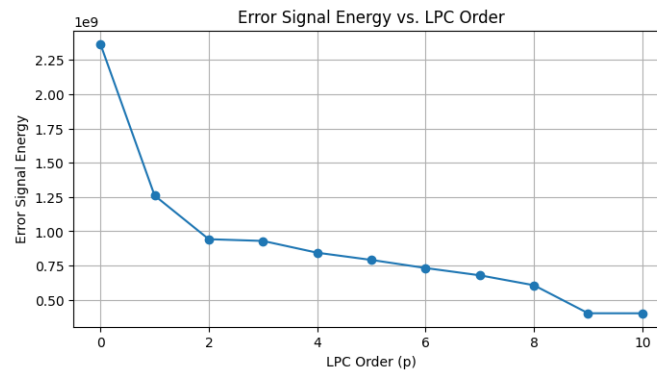
4.1 Solution 3

Using the recursion algorithm that was seen in the class, computed the LP coefficients for each order p.

4.2 Code

The code files are included in the .ipynb file included in the submission.

4.3   Plots

4.4   Observations

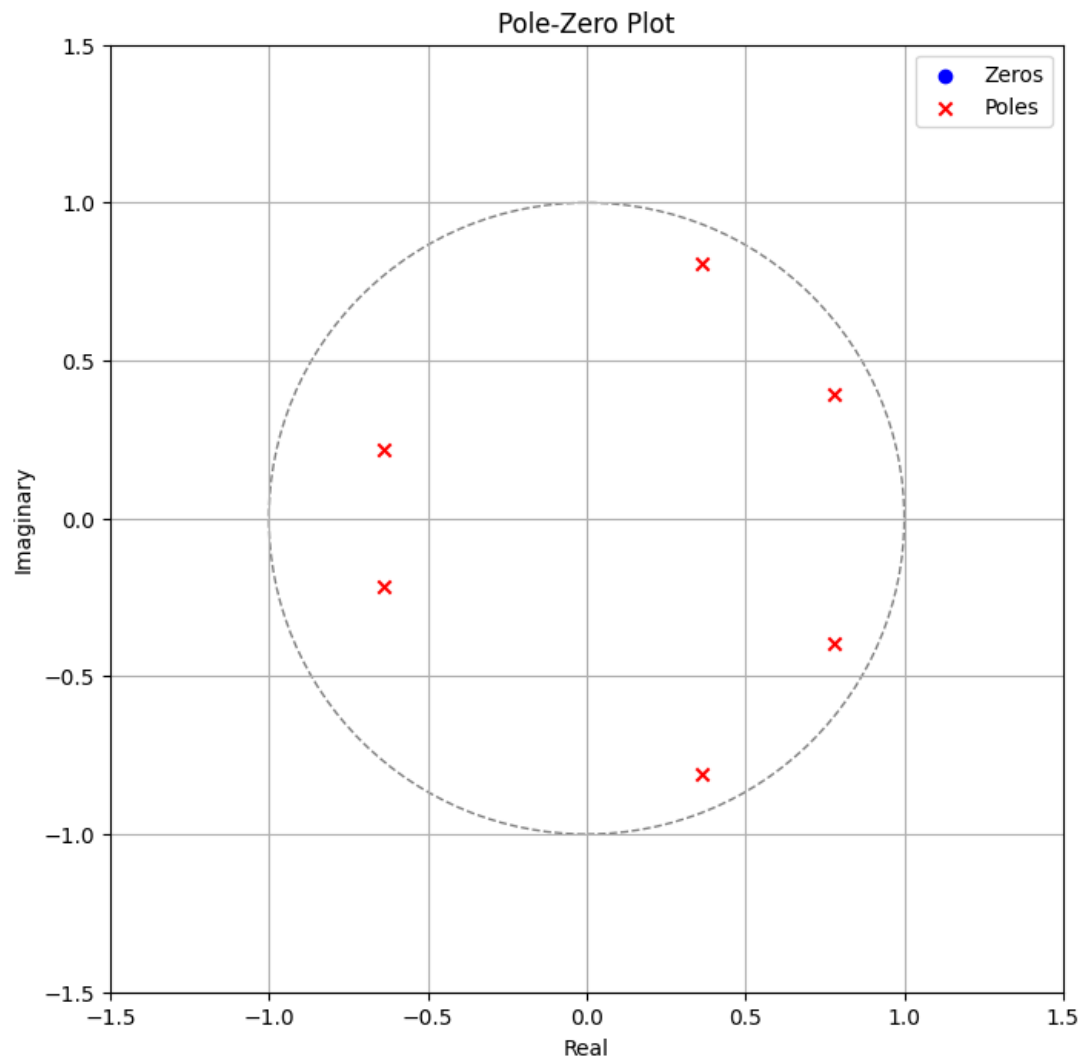We observe that the error continuously decreases as we increase the value of p which was expected.
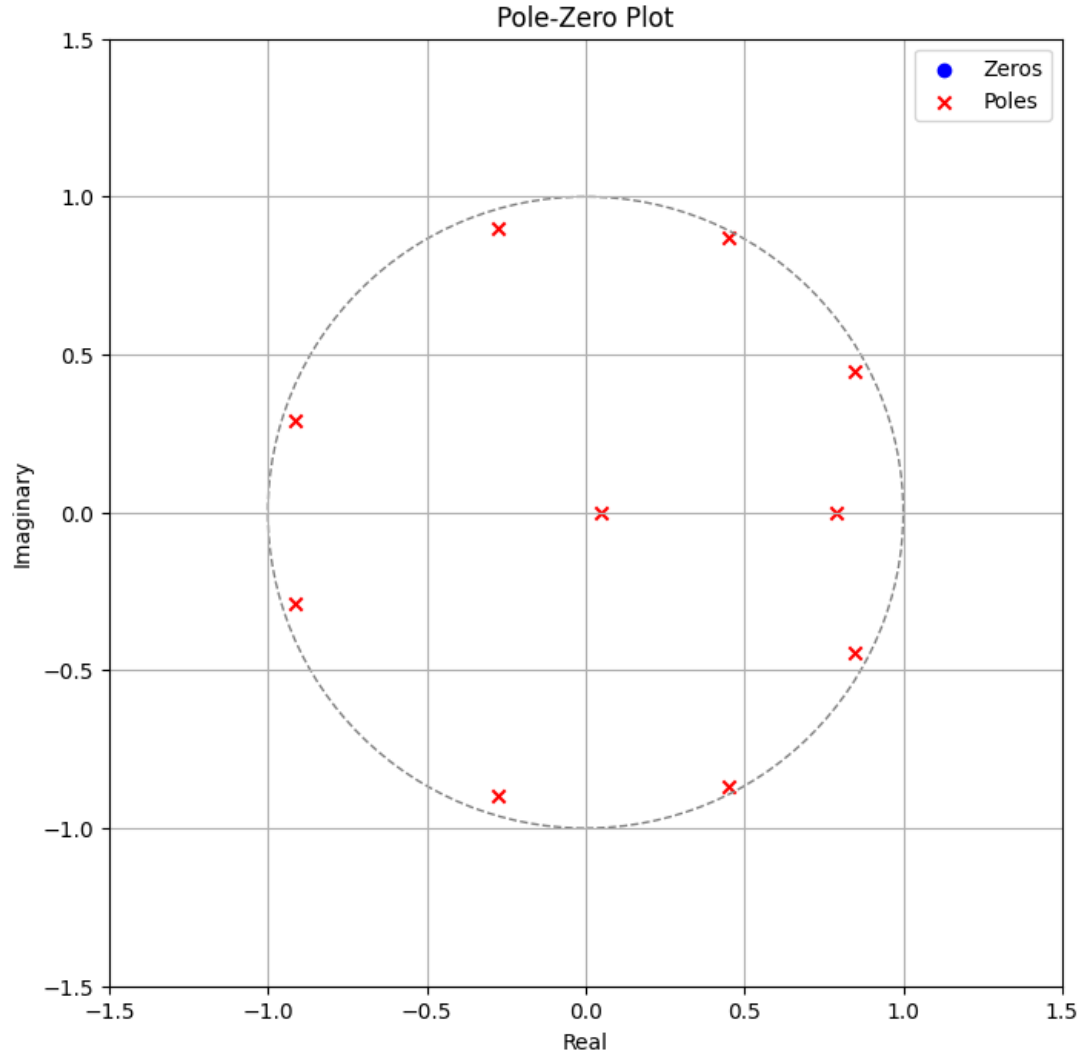
5   Question 4

Show the pole-zero plots of the estimated all-pole filter for p=6,10; comment.

5.1   Code

The code files are included in the .ipynb file included in the submission.

5.2   Plots

5.3 Observations

- For p = 6, we get 6 poles in the pole zero plot. In order to recalculate the formant frequencies and bandwidths of these formants, it has identified three formants whose phase and magnitude can be used. Since the z-plane poles' radius is farther from the unit circle in this situation, their bandwidth spread would be wider.

- For p = 10, we get 10 poles in the pole zero plot. Therefore we observe that the poles are nearly coinciding on the unit circle meaning that we can more accurately calculate the actual bandwidths.

6 Question 5

Compute the gain and plot the LPC spectrum magnitude (i.e. the dB magnitude frequency response of the estimated all-pole filter) for each order "p". Comment on the characteristics of the spectral envelope estimates. Comment on their shapes with reference to the short-time magnitude spectrum computed in part 2.
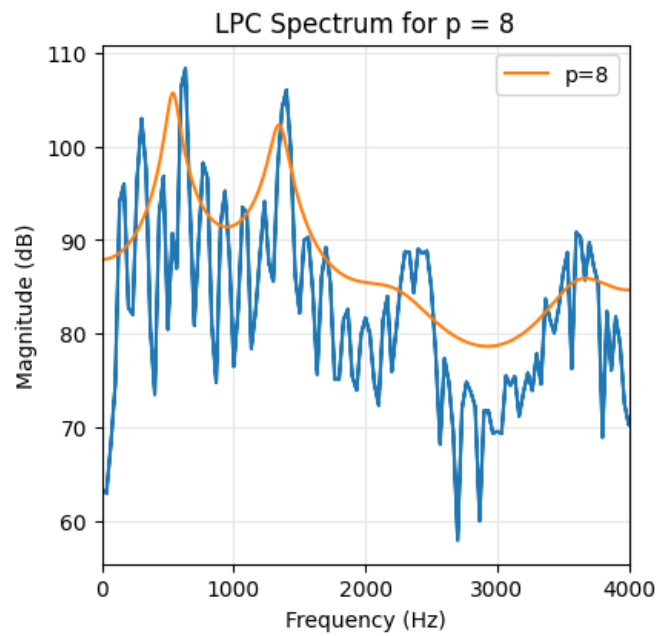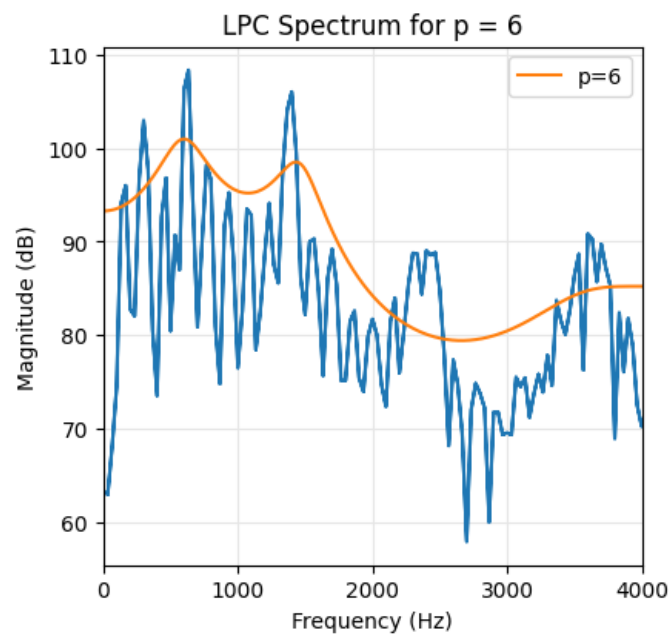
6.1 Code

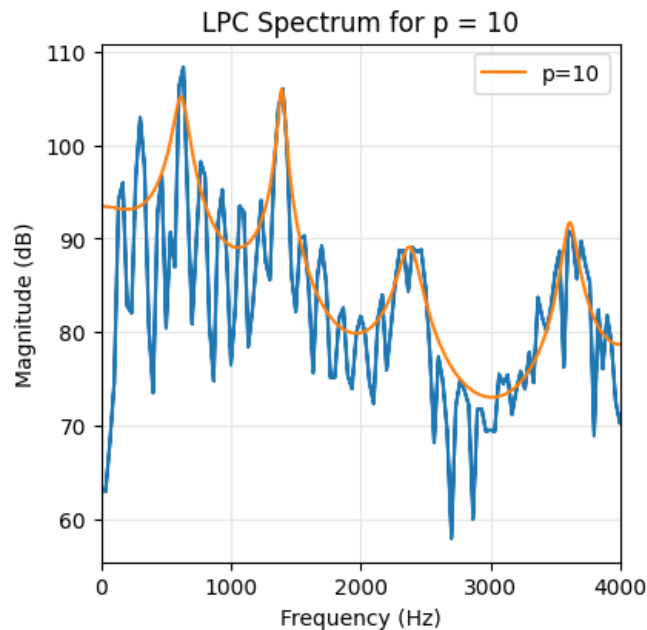The code files are included in the .ipynb file included in the submission.

6.2 Plots

LPC Spectrum for p = 2

LPC Spectrum for p = 4

LPC Spectrum for p = 6

LPC Spectrum for p = 8

LPC Spectrum for p = 10
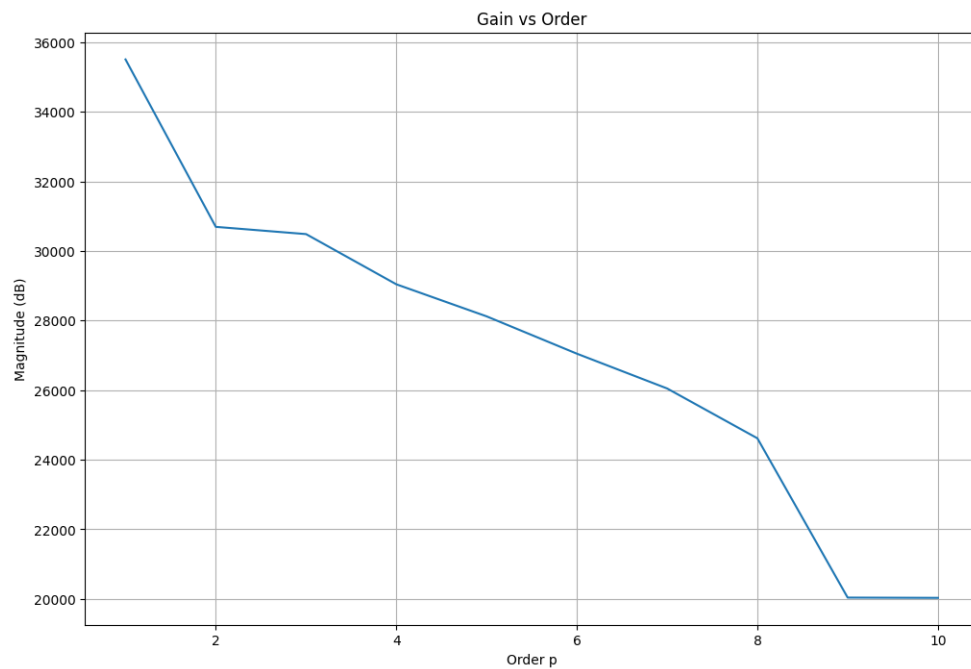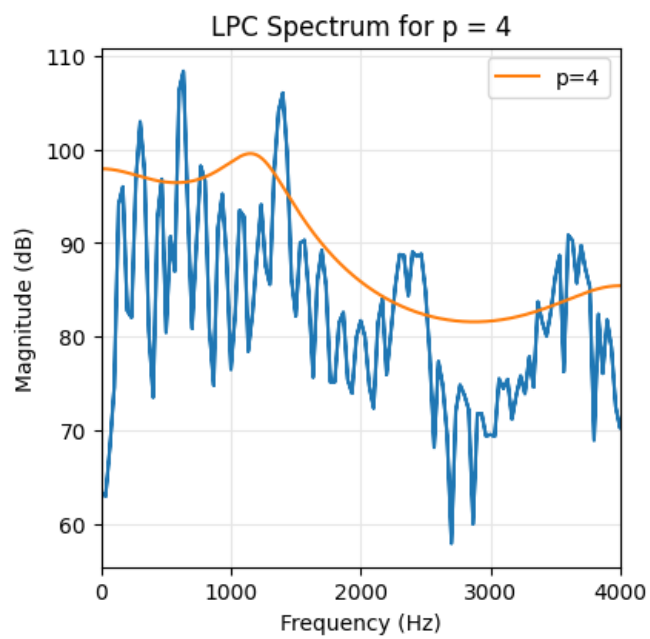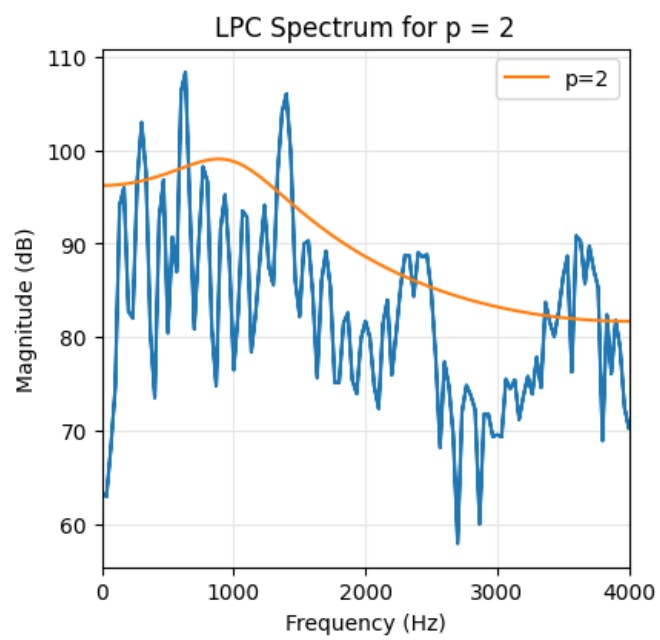
6.3 Observations

- We can easily observe that as the order of LPC coefficients increase we are more accurately reaching the actual magnitude spectrum.

- Also we observe that the gain keeps on decreasing with increasing order because it is directly proportional to the square root of energy so since energy decreases continuously the gain also decreases.

- Number of formants also increase as we increase the order p.

7 Question 6

Based on the 10th-order LP coefficients, carry out the inverse filtering of the /a/ vowel segment to obtain the residual error signal. Can you measure the pitch period of the voiced sound from the residual waveform? Use the acf to detect the pitch. Compare the acf plots of the original speech and residual signals.
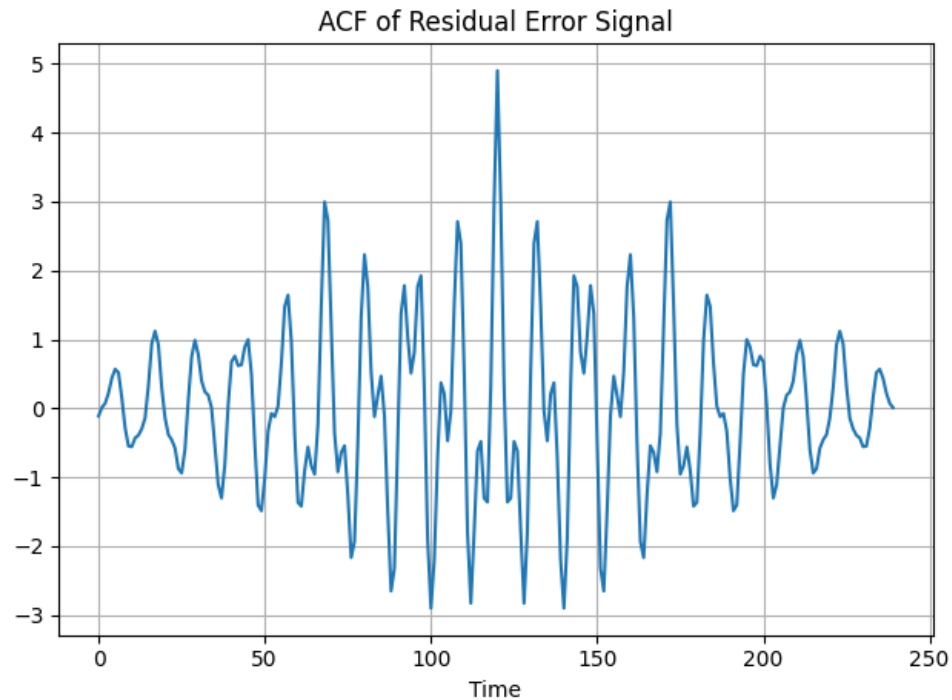
7.1 Code

The code files are included in the .ipynb file included in the submission.

7.2 Plots

Residual Error Signal

ACF of Original Sound

ACF of Residual Error Signal
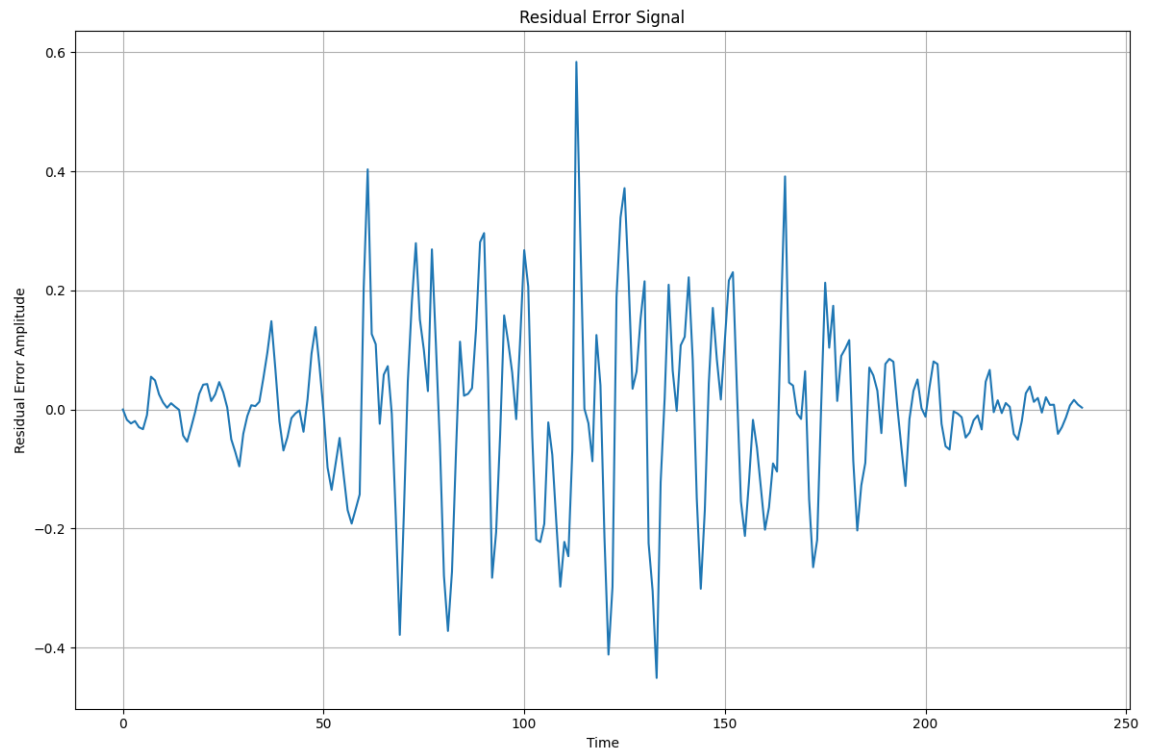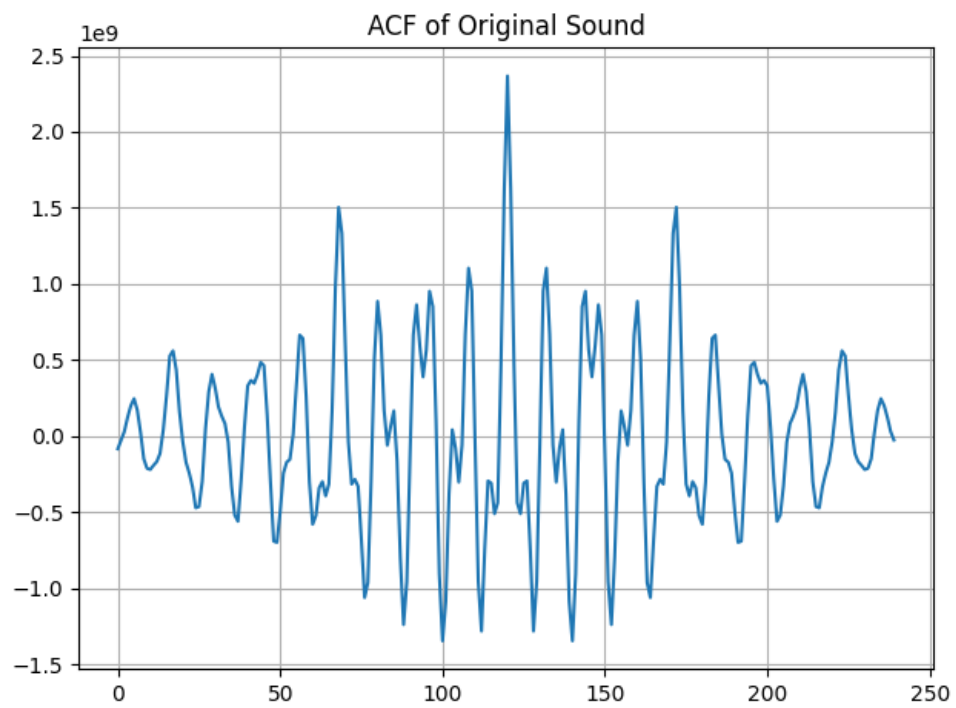
7.3 Observations

- Yes, we can measure the pitch period of the voiced sound from the residual signal as shown in the code by using the distance between 2 consecutive maximas.

- Also the shape of Original and Residual Signal is almost same except the magnitude which means that the actuala nd predicted signal is almost same.

- The pitch calculated by the code in this assignment is 153 Hz approx.

8 Question 7

LP re-synthesis: We analysed a natural speech sound /uh/ above. Using a suitable set of parameter estimates as obtained there, we wish to reconstruct the sound.

That is, use the best estimated LP filter with an ideal impulse train of the estimated pitch period as source excitation. Carry out de-emphasis on the output waveform. Set the duration of the synthesized sound to be 300 ms at 8
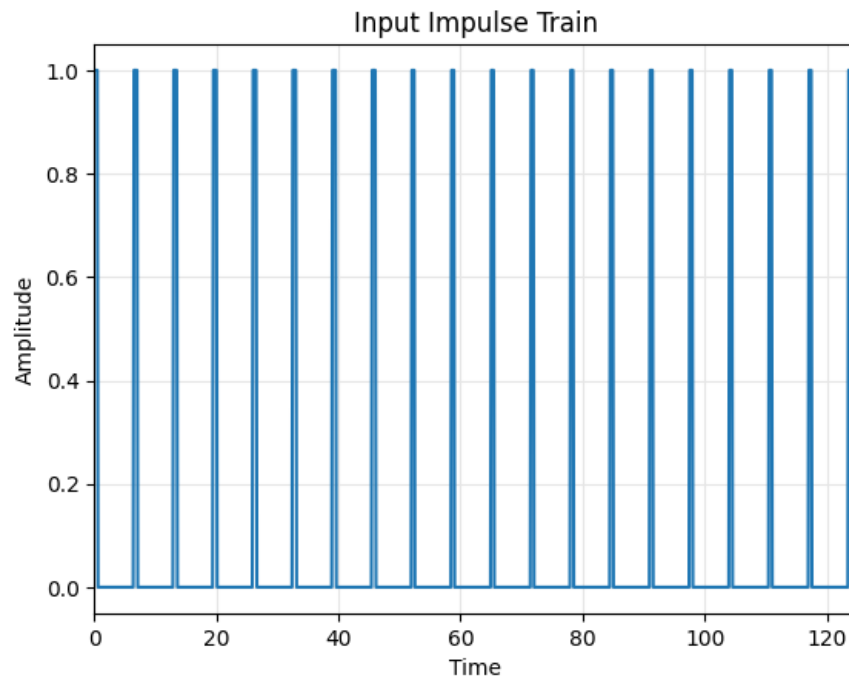
kHz sampling frequency and view the waveform as well as listen to your created sound.
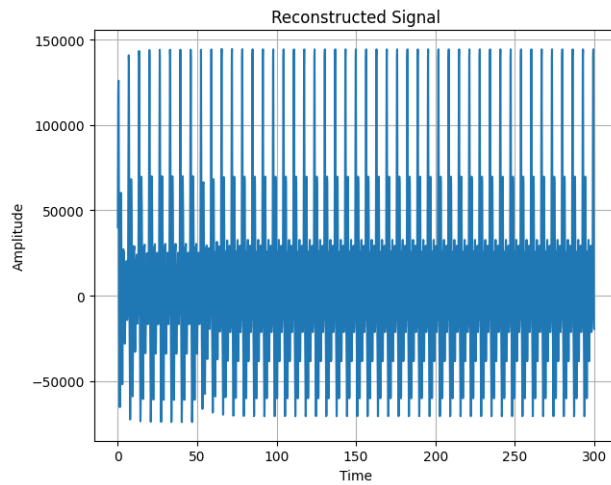
Comment on the similarity with the original sound. Try out voice modification using this analysis-synthesis method (e.g. change the voice pitch).
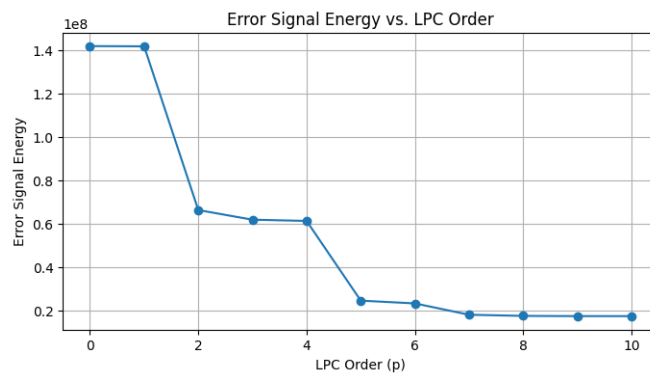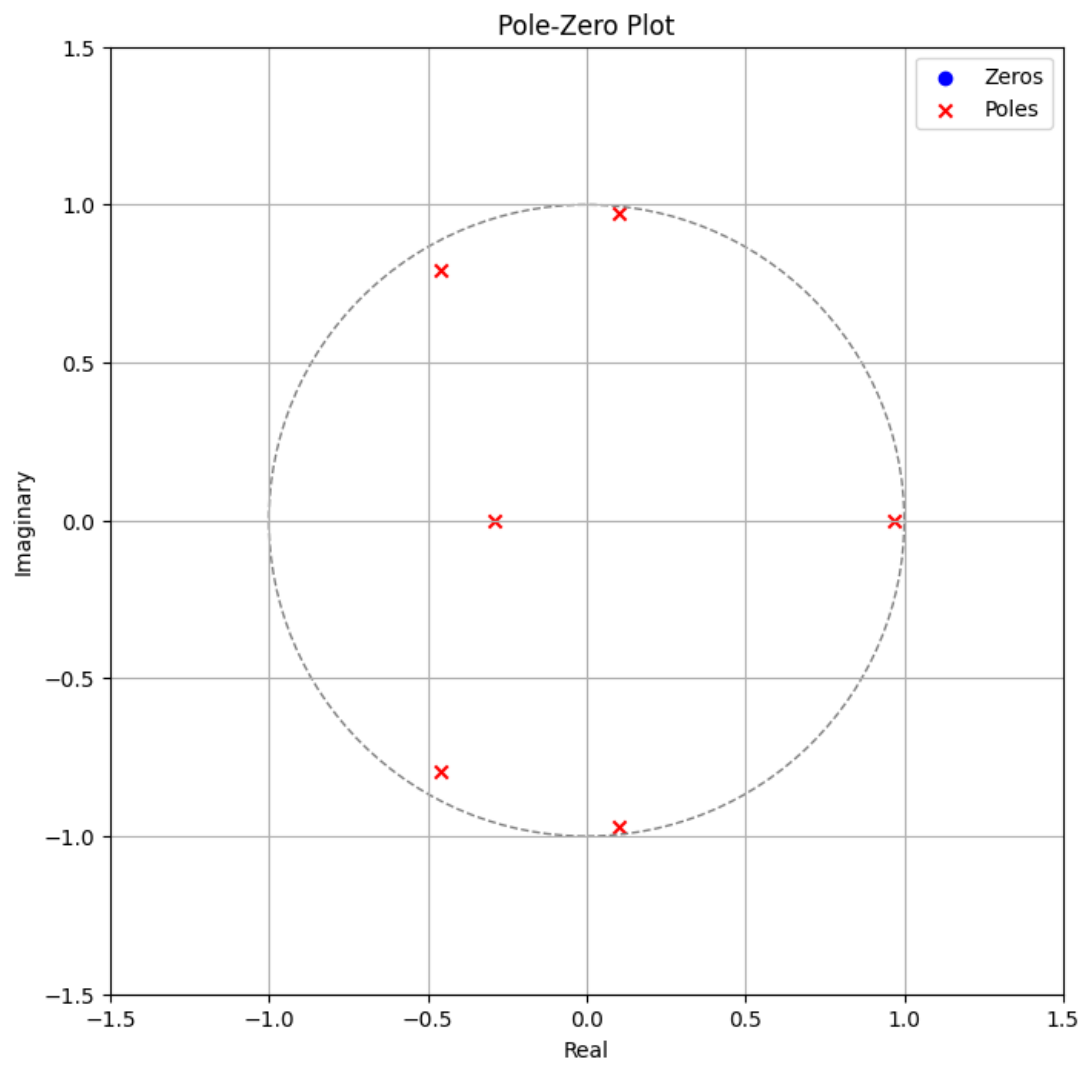
8.1 Code

The code files are included in the .ipynb file included in the submission.

8.2 Plots

Reconstructed Signal

8.3 Observations

- After calculating the impulse train we apply that on the signal and then apply de-emphasis because we were working on the emphasis signal from start.

- The reformed sound is almost similar to the actual sound.

9 Question 8 (BONUS)

Perform LP analysis on the provided /s/ sound (sampled at 16 kHz).

9.1 Code

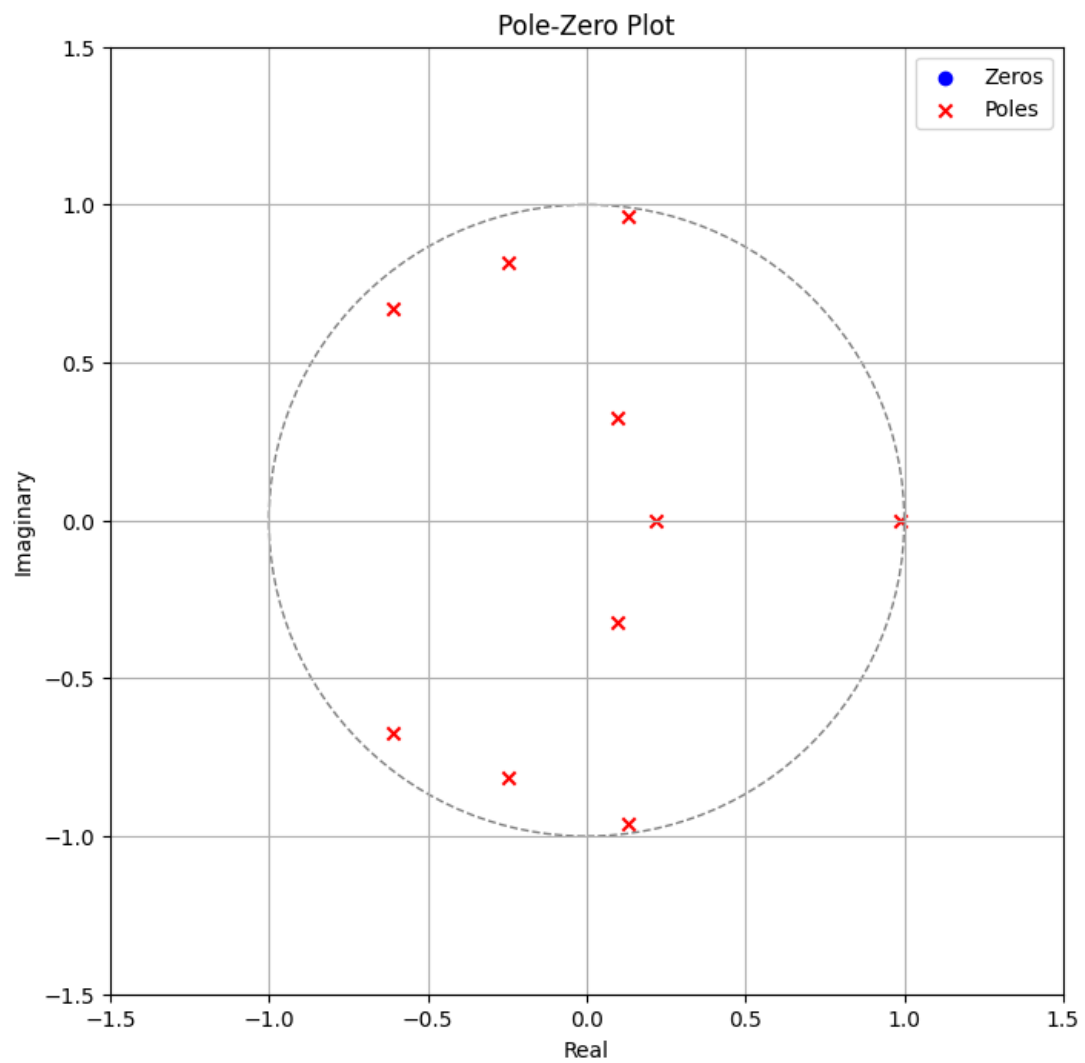The code files are included in the .ipynb file included in the submission.

9.2 Plots

Actual Signal Plot

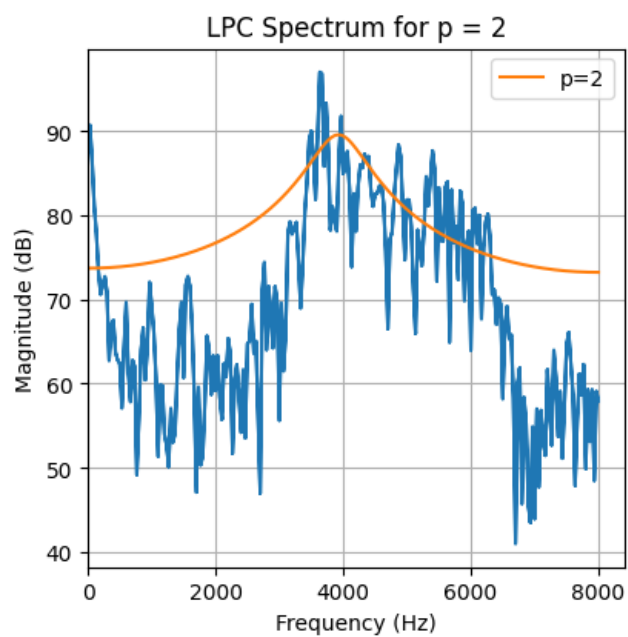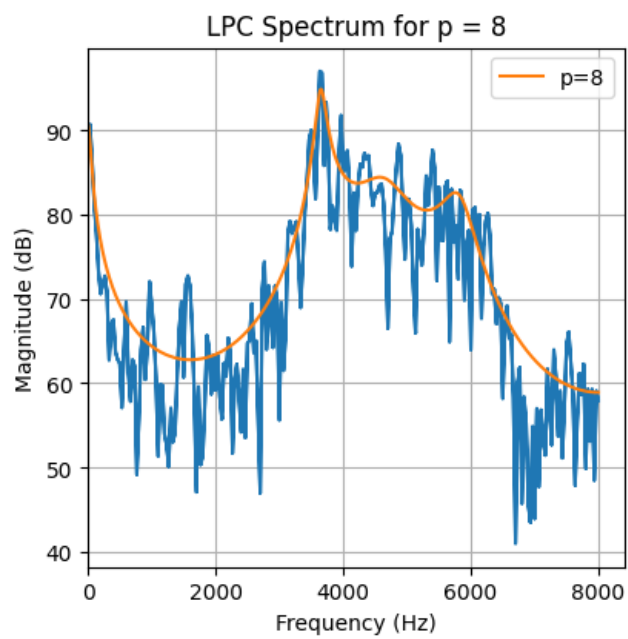(plot: Amplitude vs Time, y-axis from -4000 to 4000, x-axis from 0 to 1400)

Emphasized Signal Plot

(plot: Amplitude vs Time, y-axis from -4000 to 6000, x-axis from 0 to 1400)

Narrowband Magnitude Spectrum

(plot: Magnitude (dB) vs Frequency (Hz), y-axis from 40 to 90, x-axis from 0 to 8000)

Error Signal Energy vs. LPC Order

Pole-Zero Plot

Pole-Zero Plot

LPC Spectrum for p = 2

LPC Spectrum for p = 4

LPC Spectrum for p = 6

LPC Spectrum for p = 8

LPC Spectrum for p = 10

Gain vs Order

Residual Error Signal

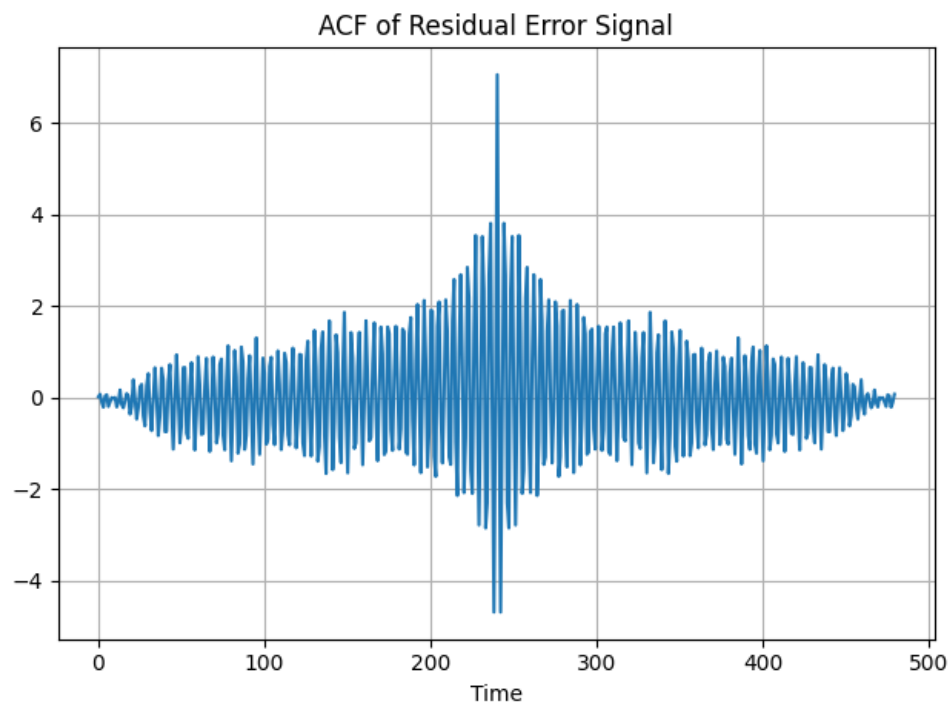ACF of Residual Error Signal
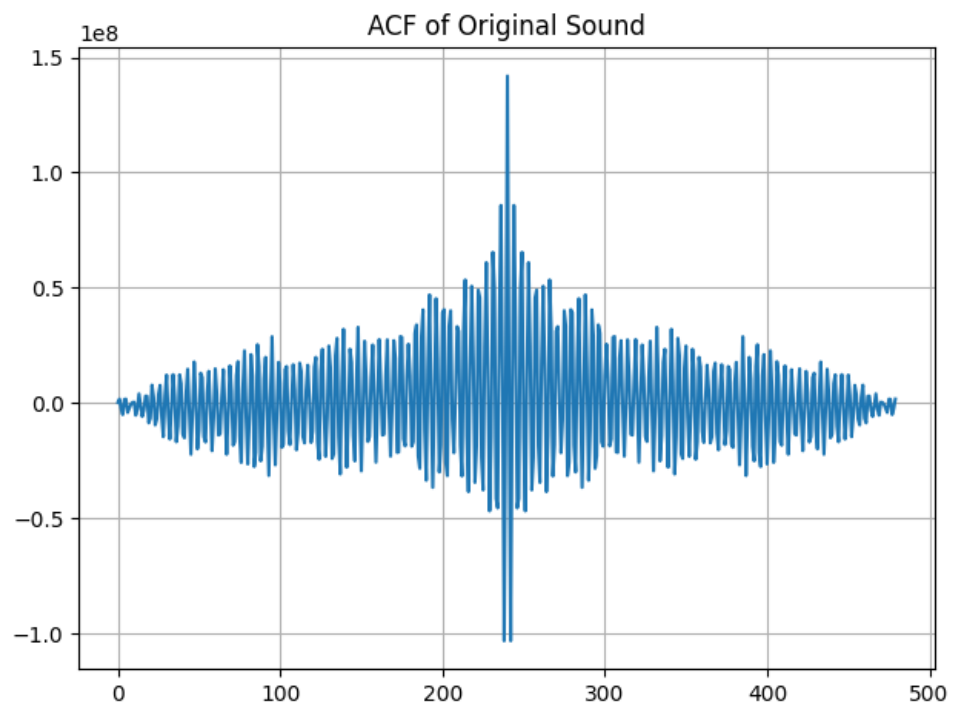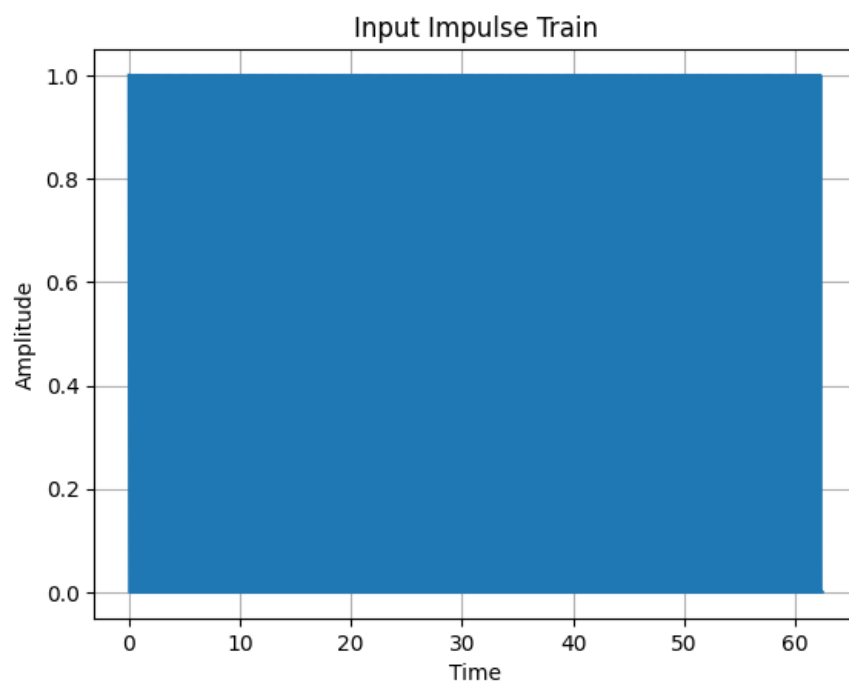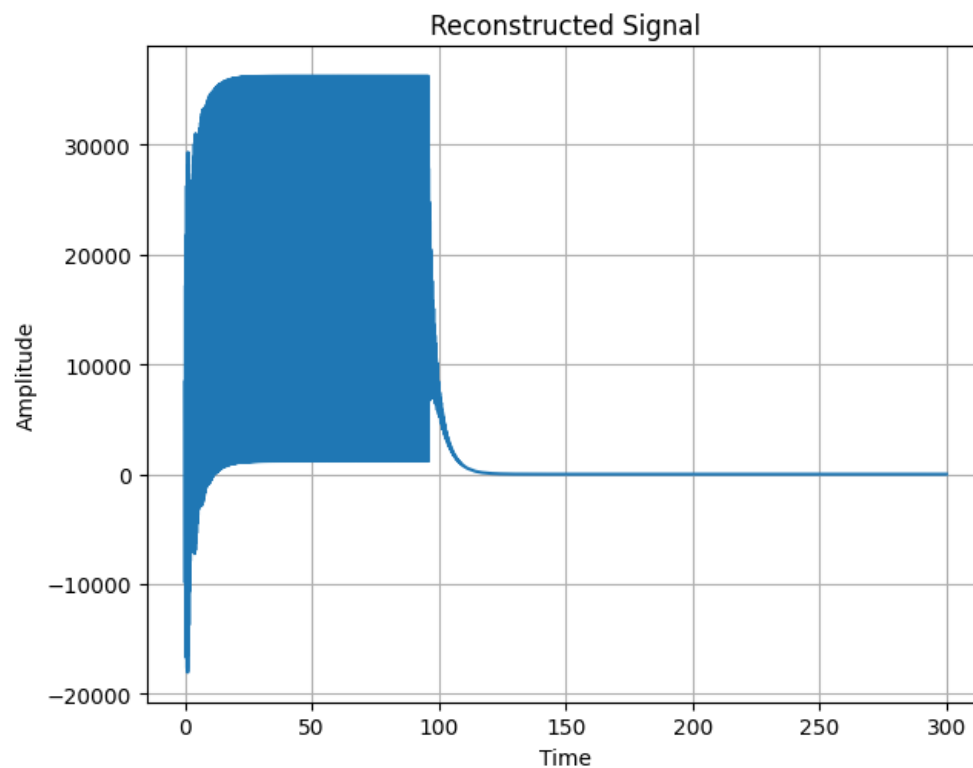
ACF of Original Sound

Input Impulse Train

9.3   Observations

Now we apply all the things we have done in this assignment to a new audio signal which is given here and the results can be seen in the notebook. We can't calculate the pitch accurately from the second signal i.e. the s sound because it is an unvoiced signal