CS6300 Assignment 3 Report

Sandeep Kumar Suresh - CS22Z121, Praveen S V - CS21D201

17 February, 2023

1 Task 1

We cannot take the whole signal as a whole and compute the LPC parameters since speech is a time varying signal.Therefore we are considering a small window where the speech property are not changing.

We can use LPC coefiicient as a feature for reconstructing the speech signal.This is expected to give convey the information as the orginal signal , but will have a bad quality.

Considering the Vocal Tract as a System , which is linear and not time-varying , LP Coefficients model's the vocal tract.

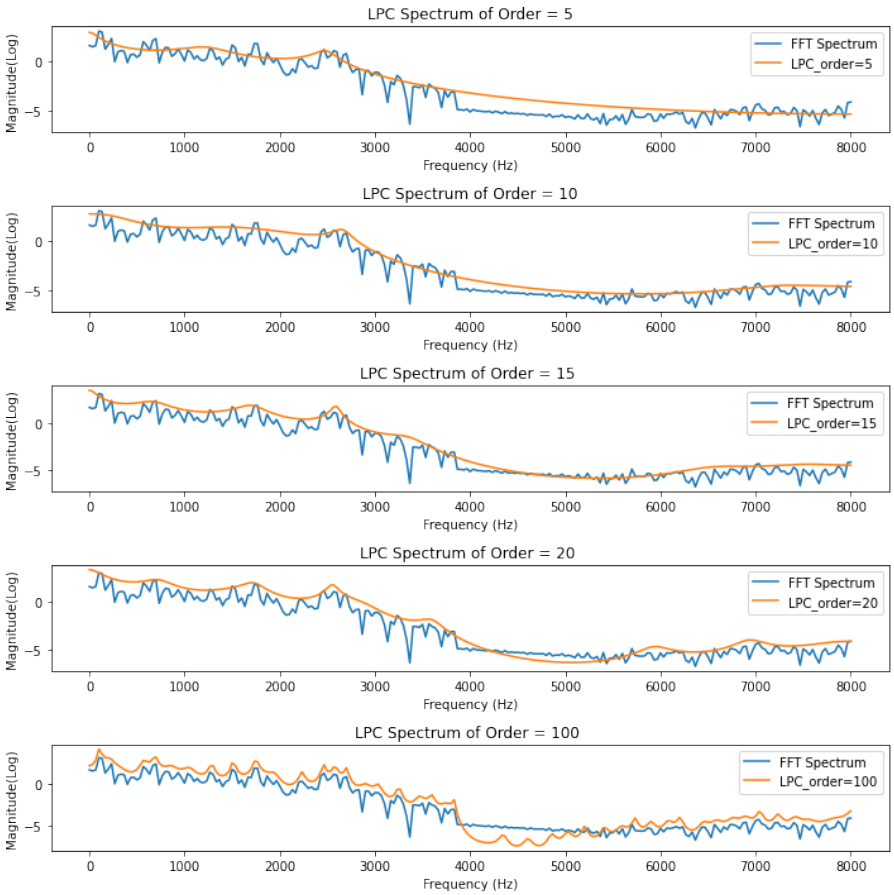

Figure 1: LPC Spectrum of Different Order

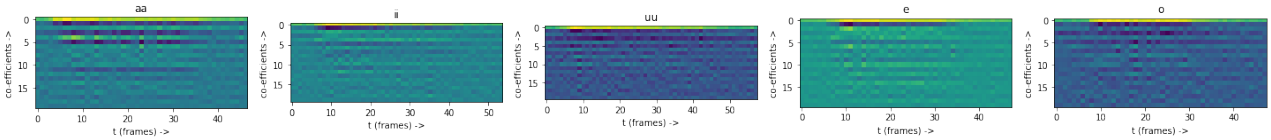

Figure 2: Visualizing co-efficients

1.1 Observations

- As we can see from the different order of LPC Coefficients, when there is a sudden change in the magnitude of the spectrum , lower order LPC Coefficients are not able to follow the envelope of the DFT Spectrum
- If we take an higher order spectrum like P = 100 where P is the order of LPC Coefficients , then the LPC Spectrum envelope the DFT Spectrum completely.
- As P increases the LPC Spectrum matches the DFT Spectrum . We know that the LPC Coefficients model the Vocal Tract, so therefore, we need to select a P value that can approximate the DFT Spectrum while preserving all the information. Here P = 15 preserves most of the information of the spectrum.
- LPC is like a Compression Coefficient of DFT. It might be possible to synthesis the sound back using the LPC Coefficients.
- The co-efficients, as visualized in Figure 2 help us understand that sound signals are not completely stationary.
- However, zooming in to Figure 2 and observing the small windows, we notice certain repeating patterns, which shows the benefit of windowing signals to get it to be quasi-stationary.

2 Task 2

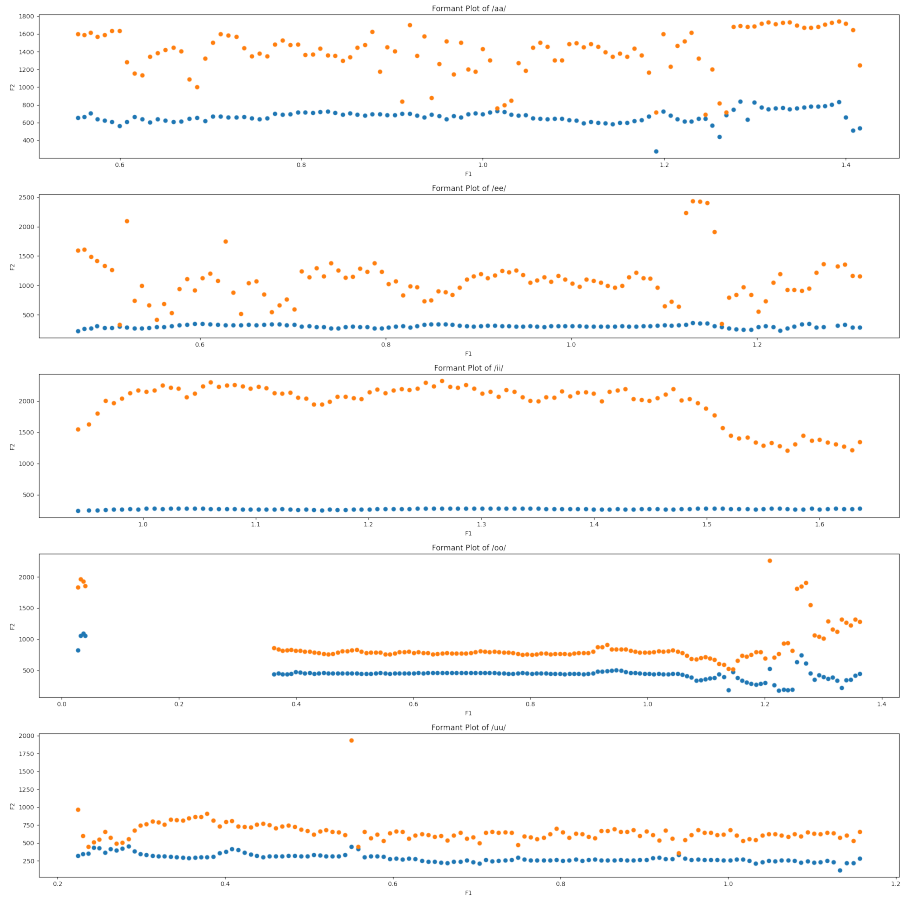

Figure 3: Formant Contours

- The formant contours closely resemble the ones we obtain in Assignment 2 using Praat.
- It seems possible to distinguish between vowels just by using the F1 and F2 values.

3 Task 3

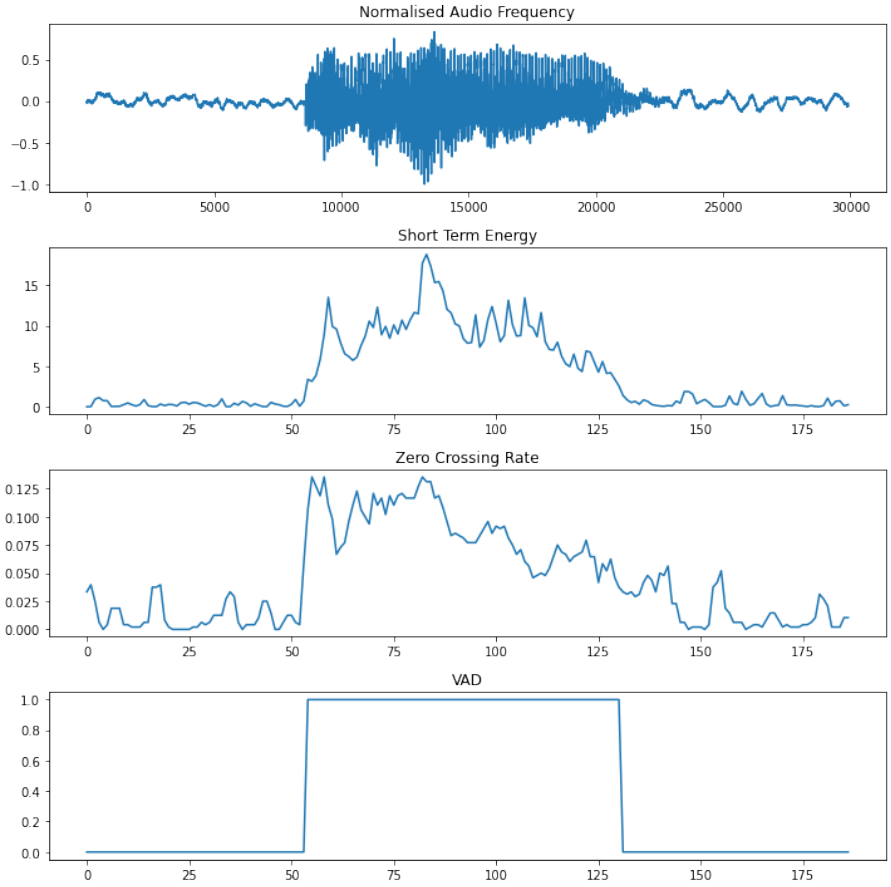

Figure 4: VAD

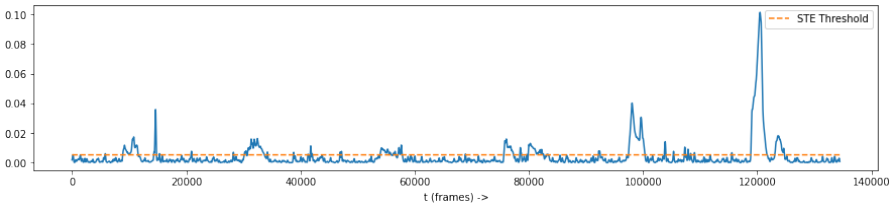

Figure 5: STE for speech with pauses in noisy background

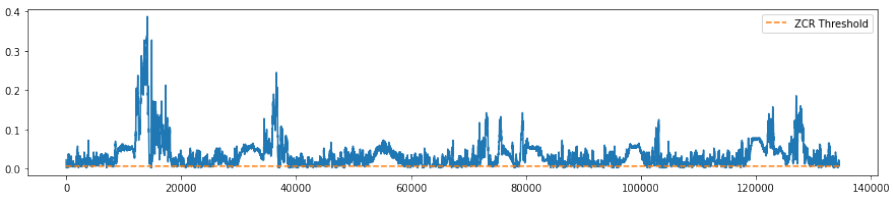

Figure 6: ZCR for speech with pauses in noisy background

3.1 Observations

- As observed in Figure 4, in the Unvoiced Region , due to noise the ZCR(Zero Crossing Rate) is high compared to the Voiced Region.The ZCR in Voiced region could be due to the periodic of the vowel caused due to the vocal ttract vibration.

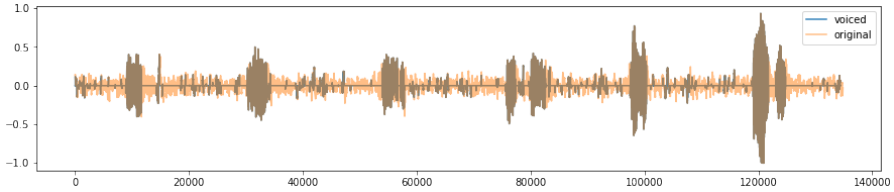

Figure 7: Voiced vs. Unvoiced signal for speech with pauses in noisy background

- In STE (Short Term Energy) Analysis The Voiced Region has higher energy compared to the unvoiced Region.
- For VAD(Voice Activity Detection) of Speech Signal using both STE and ZCR we need to select the region with high STE and low ZCR.
- In Figure 5, 6, 7, we visualize the short-term energies, zero-crossing rate and voiced vs. unvoiced signals for speech with pauses in noisy background, respectively. We see the method is able to detect the 6 words spoken easily while filter out a good amount of noise.
- The short-term energies clearly indicate when a word is spoken. The silence regions have background noise, and while having lower energies, the values are not zero.
- Similar, to the case for vowels, even at a word level, we do not notice very low zero-crossing rates when words are uttered.

4 Task 4

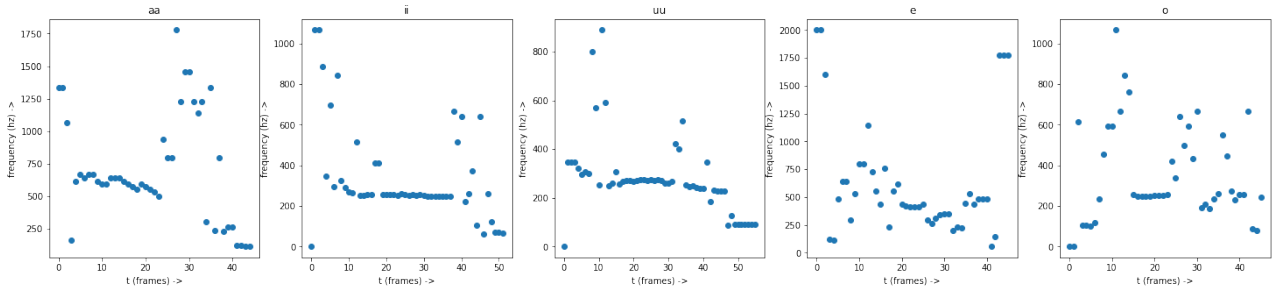

Figure 8: Pitch contours for vowels

4.1 Observations

- By performing VAD before estimating pitch, the noise regions at the before the start of the vowel and after the end of the vowel are removed to a good extent.
- For the vowel /aa/, the pitch we extract seems to match to the mean pitch estimated by Praat for the initial frames.
- For few other vowels, there is some differences in the pitch we compute and those observed from Praat in the previous assignment.
- We notice that the sometimes, the pitch contour extracted actually corresponds to a formant contour. This maybe because the formant peaks have greater magnitude than the peak corresponding to the expected pitch. This is possibly a shortcoming of using auto-correlation method for pitch or can perhaps be mitigated by using a better peak selection algorithm.

5 Task 5

5.1 Observations

- From Figure 9 we can observe that the VAD is able to detect the voiced part of speech for female voice. Similarly, pauses between words are identified by VAD as seen in Figure 11.

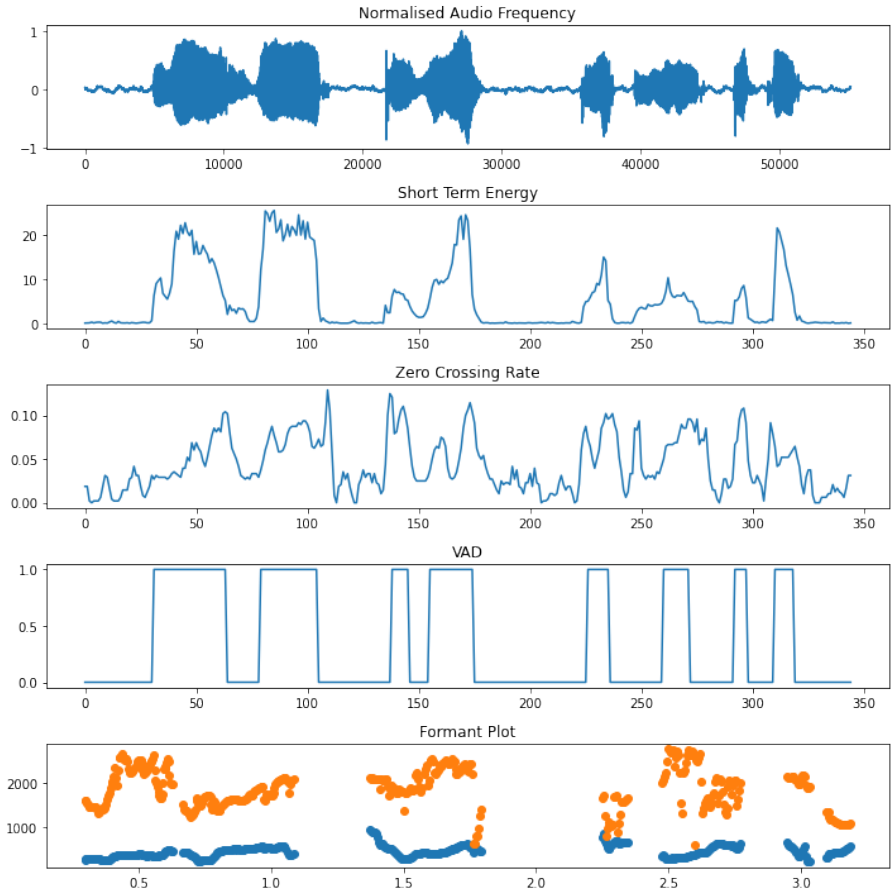

Figure 9: STE ZCR and VAD for Female Voice

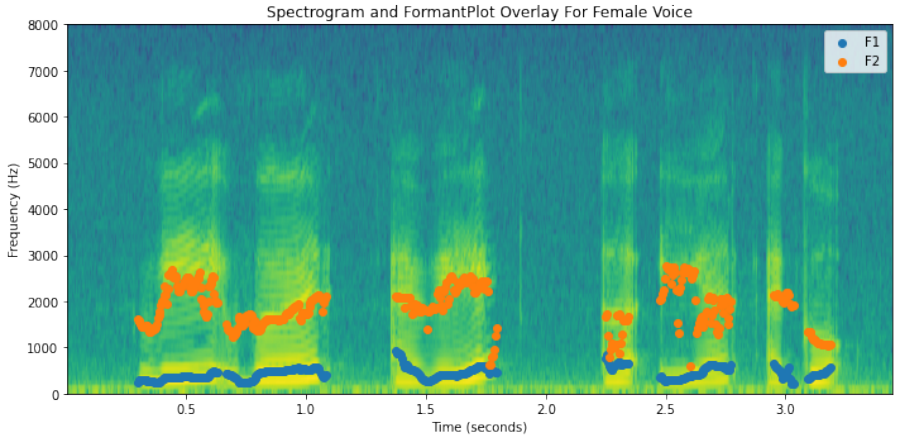

Figure 10: Spectrogram and Formant Plot Overlay

- There are more clear regions of high and low zero-crossing rates for the female speaker in comparison to that of the male speaker.
- There seems to be more variance in the formant contours of the female speaker in comparison with the male speaker.
- We observe that for both female and male speakers, the spectrogram magnitudes are significant in the region below the 7000Hz whereas higher frequencies have minimal values. This indicates that the telephone signal sampling rate of 8Khz should suffice to represent speech adequately.

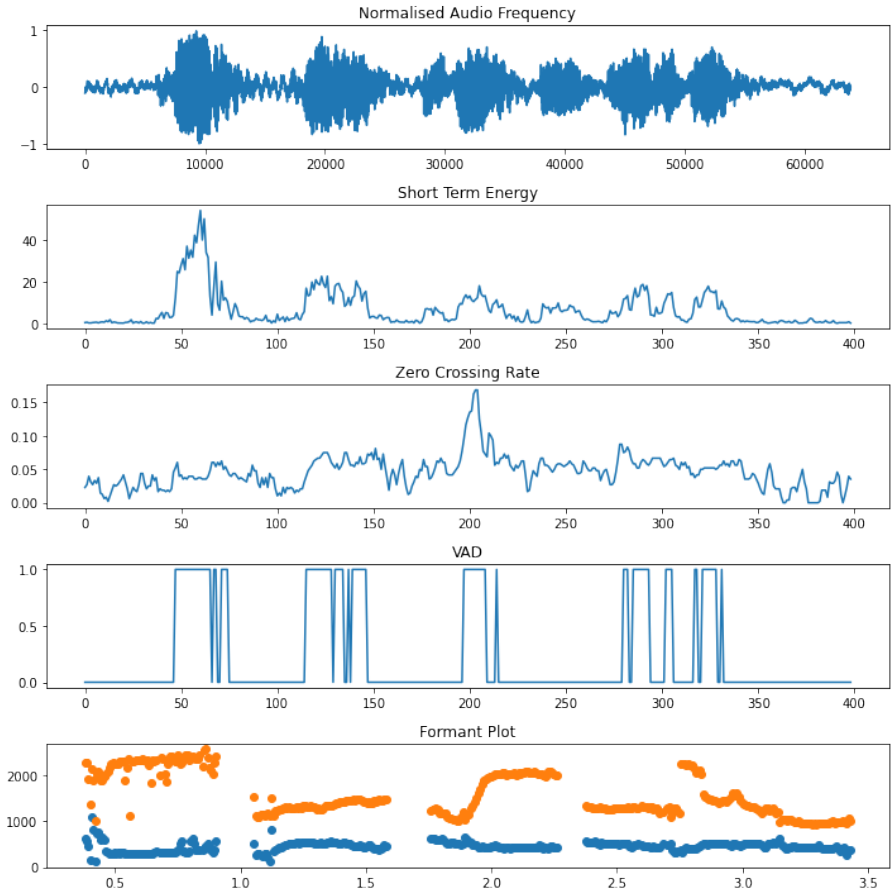

Figure 11: STE ZCR and VAD for Male Voice

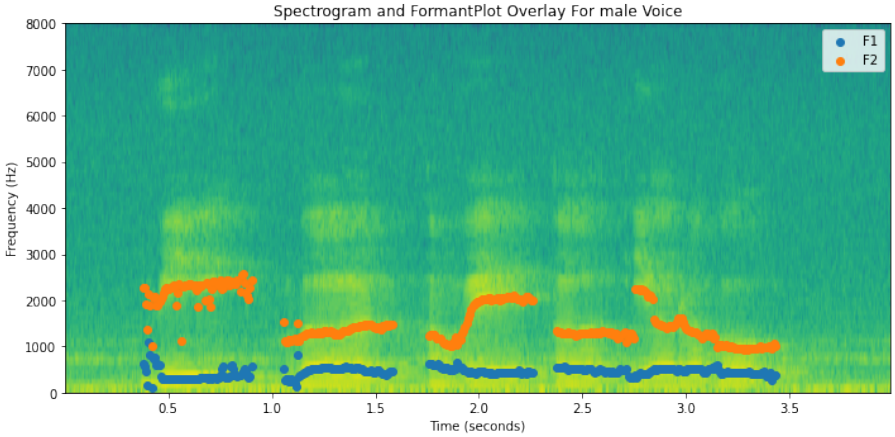

Figure 12: Spectrogram and Formant Plot Overlay -Male