Voiced/Unvoiced classification using Zero Frequency Filter

Aim

- The aim of this project is Voice/Unvoiced classification based on epoch extraction(PNZC) from speech signals using Zero Resonance Frequency Filter.
- To study and analyze the impact of source-filter interaction using ZFF method for the extraction of epochs.
- The arctic_a0001.wav is used as input speech signal.

Epoch and it’s significance

- The instant of significant excitation of the vocal-tract system is referred to as the **epoch**. Significant excitation refers to the one which is impulse-like with strength substantially larger than its neighbors.
- The source of significant excitation for voiced speech occurs at the instant of glottal closure called epoch. So speech analysis mostly depends on the accurate estimation of epoch locations within a glottal pulse.
- Knowledge of epochs helps in determining the characteristics of the voice source by a careful analysis of the signal within a glottal pulse.
- The excitation features derived from the regions around the epoch locations provide complementary speaker-specific information to the existing spectral features.

Basis for epoch extraction

- Speech is produced by the excitation of time varying vocal-tract system by 1.Glottal vibration 2.Frication 3.Burst . Glottal vibration is the primary mode.
- The excitation is significant only when there is large-energy in short time interval i.e impulse like. The presence of these impulse-like characteristics suggests that the excitation can be approximated as a sequence of impulses.
- Sequence of impulses as excitation: When an inertial system is excited by an impulse-like excitation, the effect of excitation uniformly spreads in frequency domain and is modulated by the time-varying transfer function of the system.
- The information about the time instants of occurrence of the excitation impulses reflects as discontinuities in the time domain.
- The effect of the discontinuities can be highlighted by filtering the output signal through a narrowband filter centered around a frequency.

Zero Frequency Filtering

- The discontinuities due to the excitation can be approximated by a sequence of impulses of varying amplitudes. This discontinuity is reflected across all the frequencies including zero frequency.
- The filter output of the zero frequency filter is not affected by the characteristics of time-varying vocal-tract system as it’s resonances are at high frequencies.
- The characteristics of the discontinuities can be extracted by passing the speech signal through a zero frequency filter twice. This reduces the effects of high frequency resonances.
- The output of the cascade of 2 zero frequency filters is equivalent to 4 times successive integration. Therefore the output grows/decays as a polynomial function of time.

Procedure:

- Difference the speech signal s[n] to remove any time-varying low frequency bias in the signal.

$$x[n] = s[n] - s[n-1]$$

- Obtain the l.p residual and the compute it's hilbert transform. Hilbert envelope of a signal is a positive function, giving the envelope of the signal.

$$s_h(t) = \mathrm{IFT}\left(S_h(\omega)\right)$$

- Zero frequency filtering i.e the differenced signal is twice passed through an ideal resonator at zero frequency.

$$y_1[n] = -\sum_{k=1}^{2} a_k y_1[n-k] + x[n]$$

$$y_2[n] = -\sum_{k=1}^{2} a_k y_2[n-k] + y_1[n]$$

Parameters a1= -2 and a2= 1

This is equivalent to successive integration 4 times.

- Apply the Zero frequency filtering to Hilbert envelope as well.

- Compute the mean of the window (no.of samples =2N+1).

$$\cdot \frac{1}{2N+1} \sum_{m=-N}^{N} y_2[n+m]$$

- Zero frequency filtered signal is obtained by removing the trend in Y2[n] i.e by subtracting the mean at each sample .

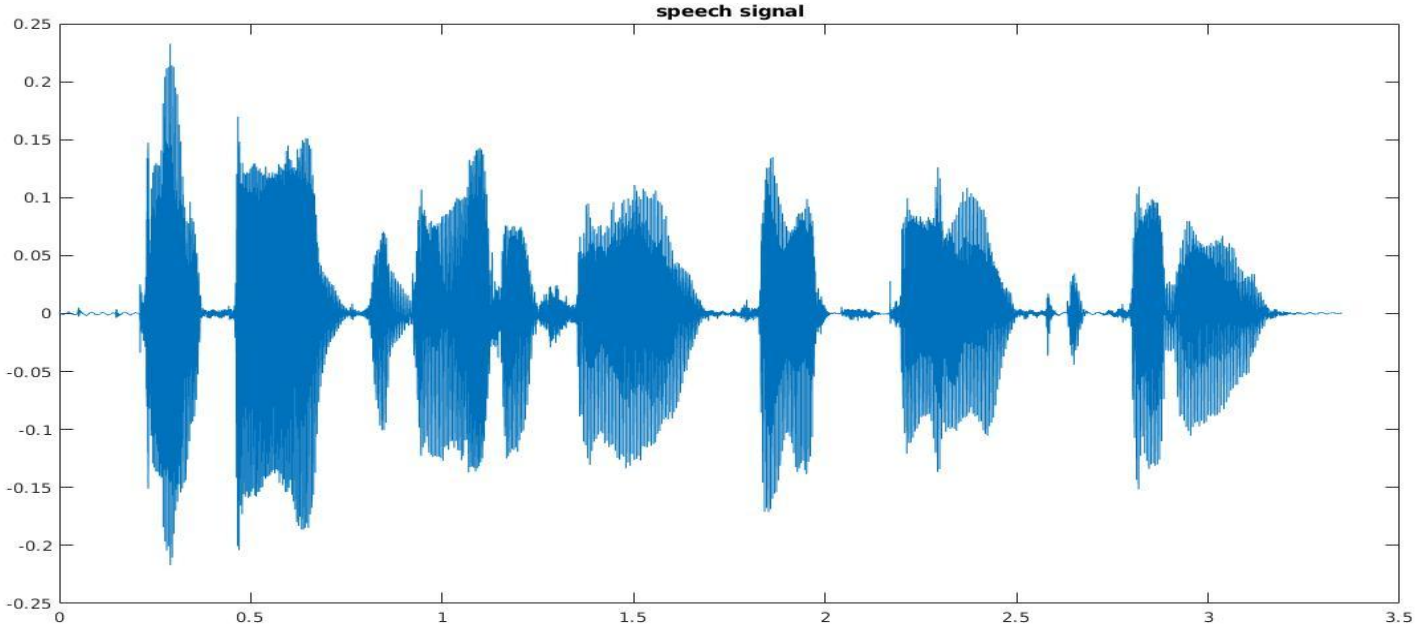
$$y[n] = y_2[n] - \frac{1}{2N+1} \sum_{m=-N}^{N} y_2[n+m]$$

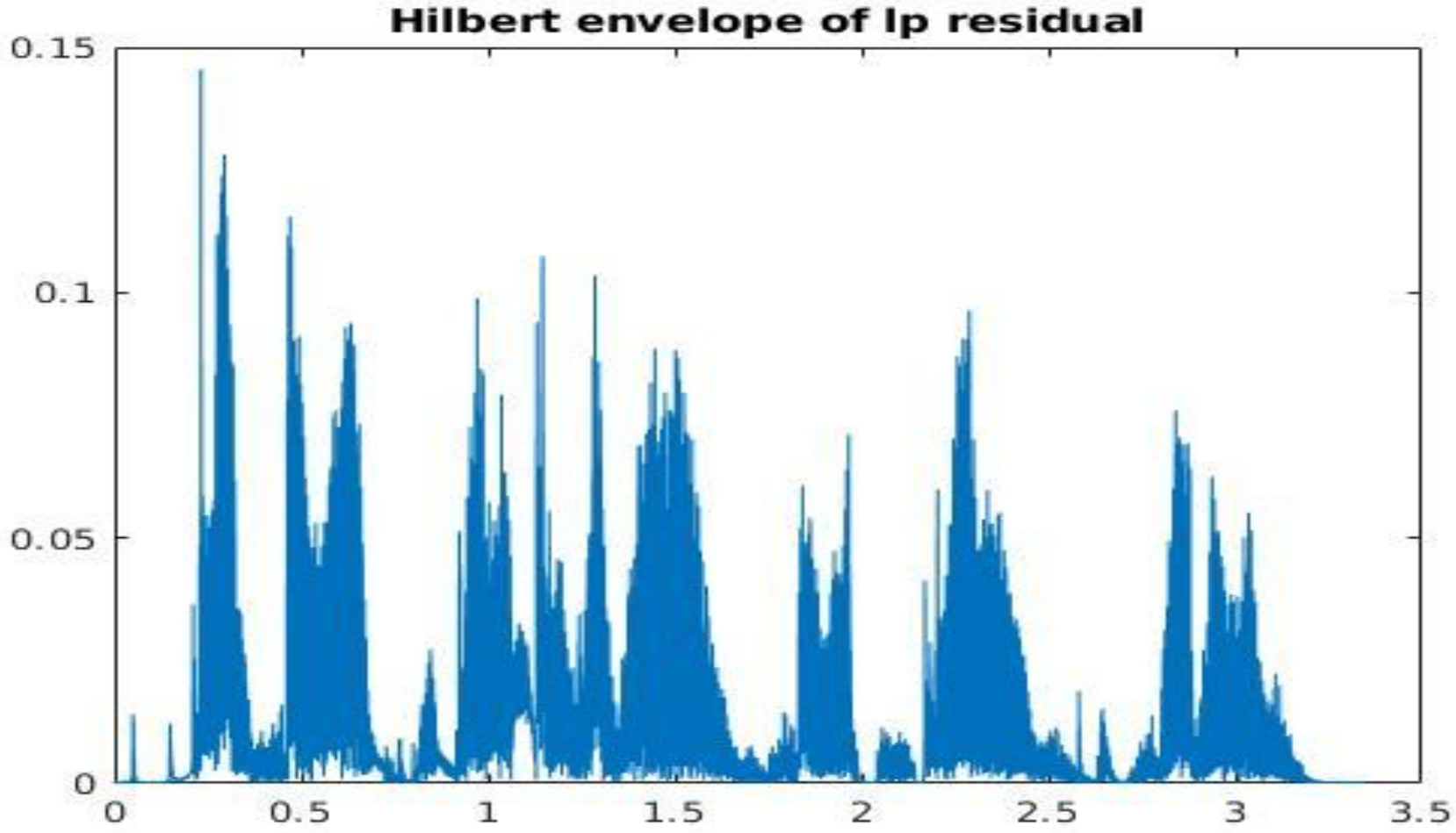
Subtract the mean from the Hilbert envelope as well in order to extract the characteristics of discontinuities.

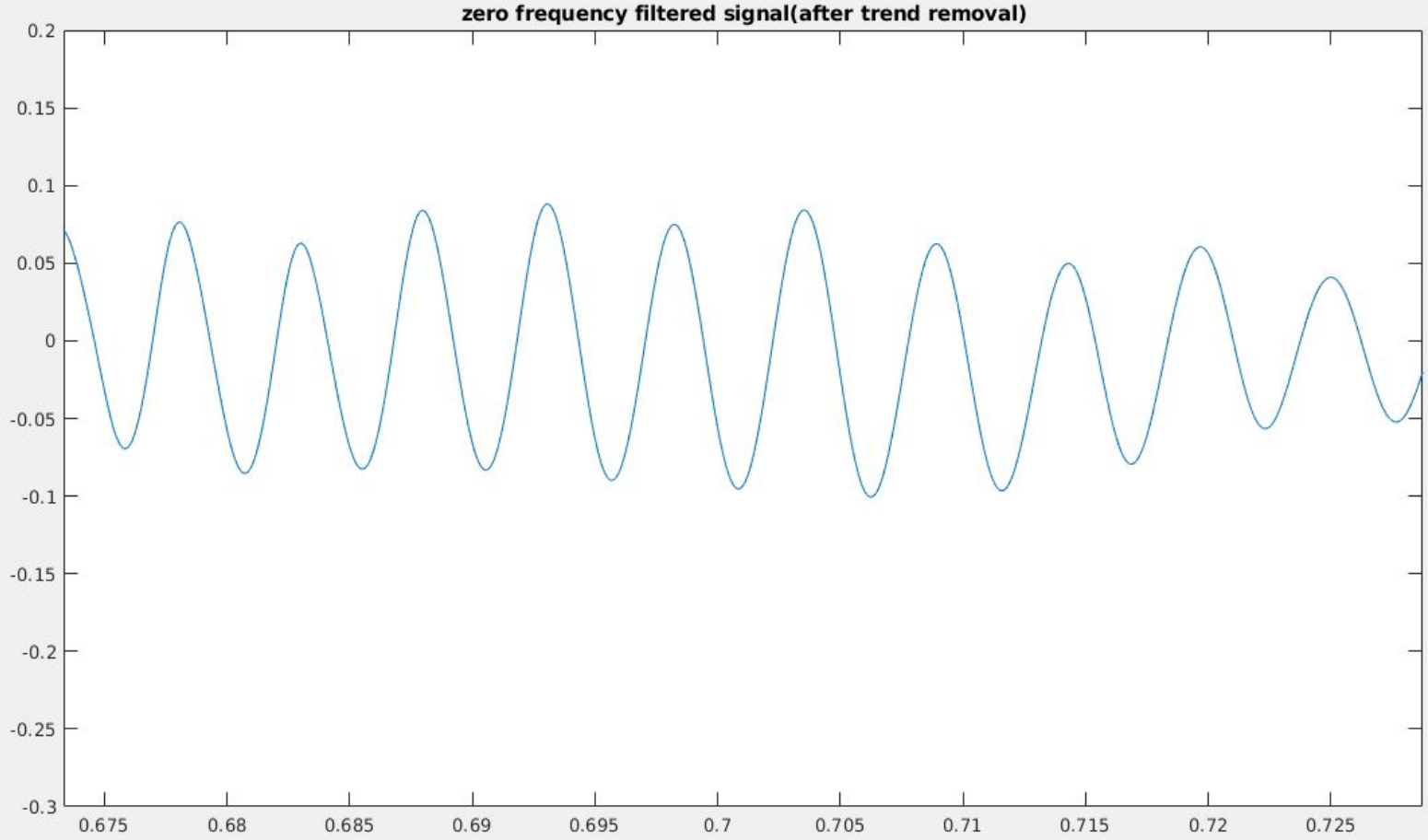
- Finally Positive to Negative Zero Crossings are the epoch locations which occur only for voiced and the remaining region is unvoiced.

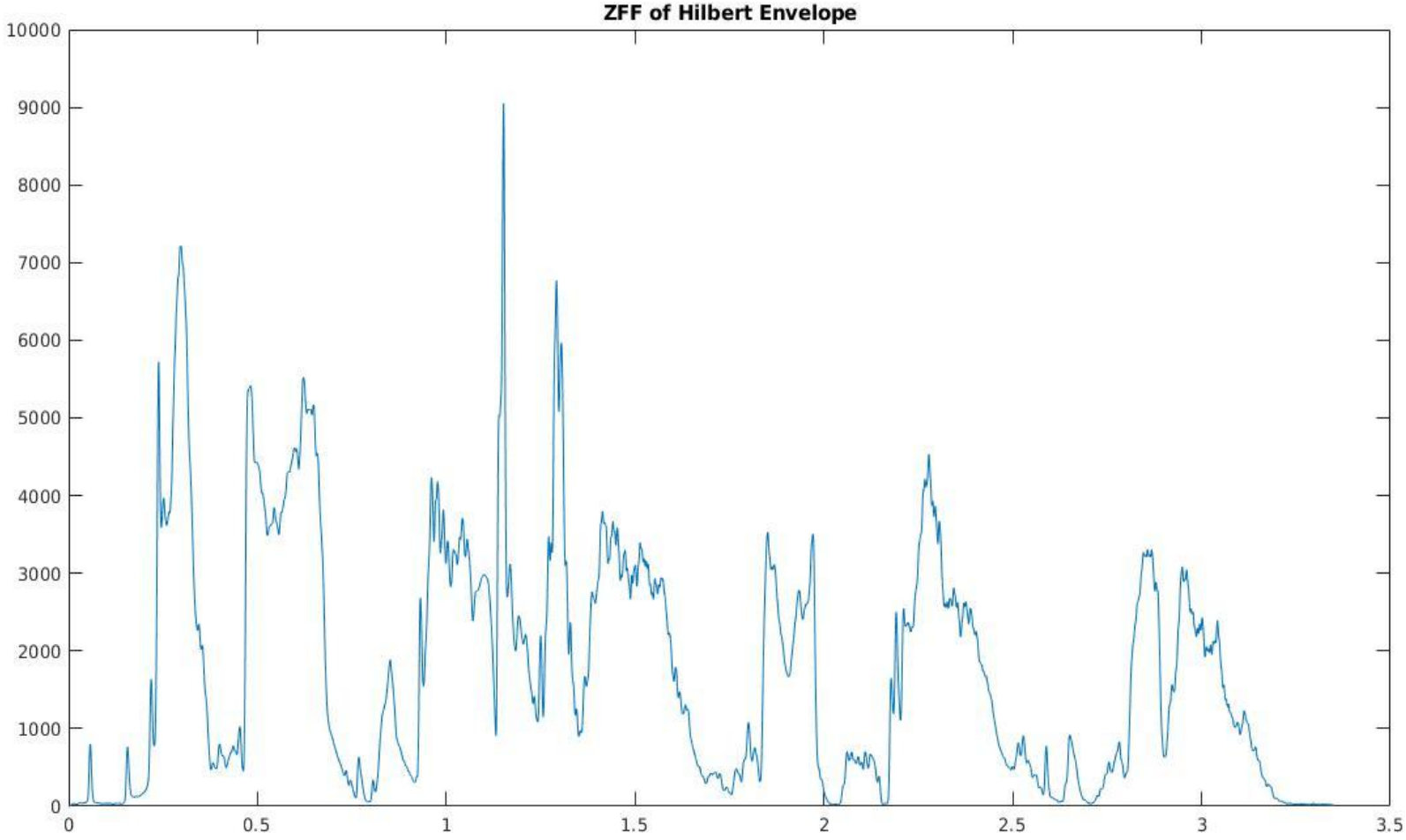
Results

Speech signal:




Hilbert envelope of lp residual
0.15
0.1
0.05
0
0
0.5
1
1.5
2
2.5
3
3.5


zero frequency filtered signal(after trend removal)
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.25
-0.3
0.675
0.68
0.685
0.69
0.695
0.7
0.705
0.71
0.715
0.72
0.725


ZFF of Hilbert Envelope
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0
0.5
1
1.5
2
2.5
3
3.5

Thank You