SPEECH SIGNAL PROCESSING

Assignment 4

Jaishnav Yarramaneni 2020102059

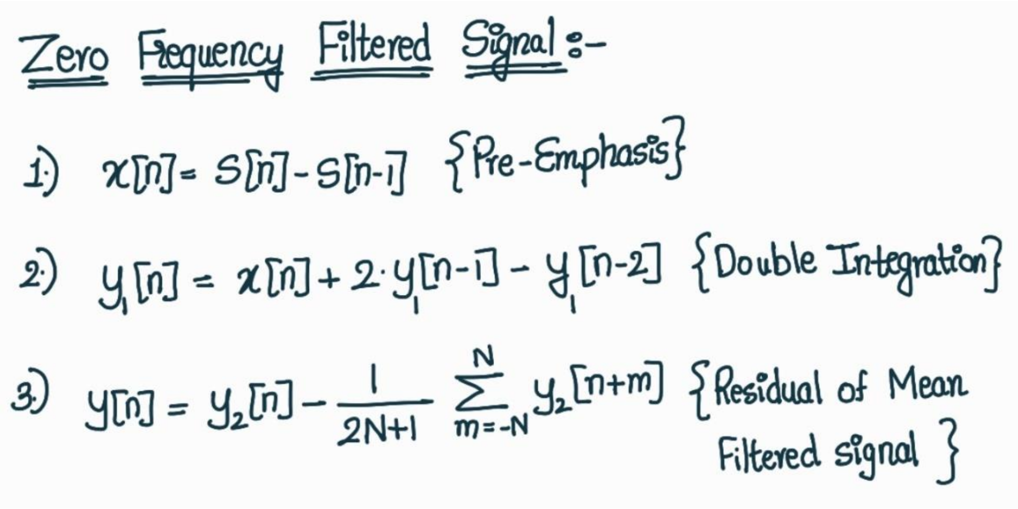
Question 1

Calculate Epochs using the ZFF approach. Note: Computer-based Question [10 points]

The epochs are found using the following steps
1) Pre-emphasis

2) Double Integration

3) Subtracting the Double Integrated signal from the mean filtered version of the double-integrated signal.



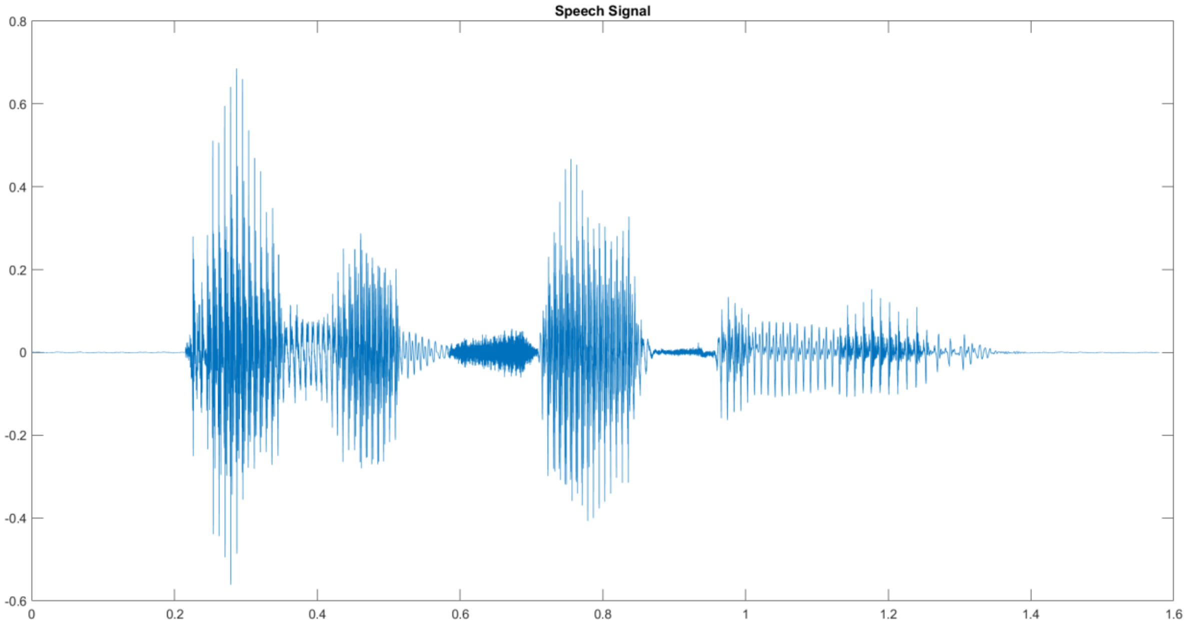
Zero Frequency Filtered Signal :-

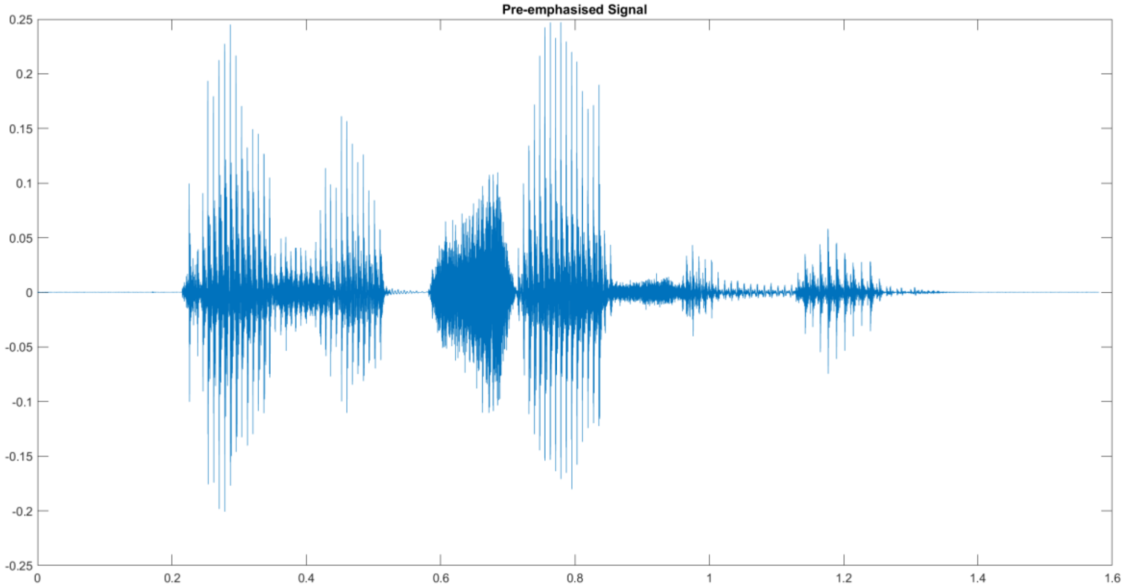
1) $x[n] = s[n] - s[n-1]$ {Pre-Emphasis}

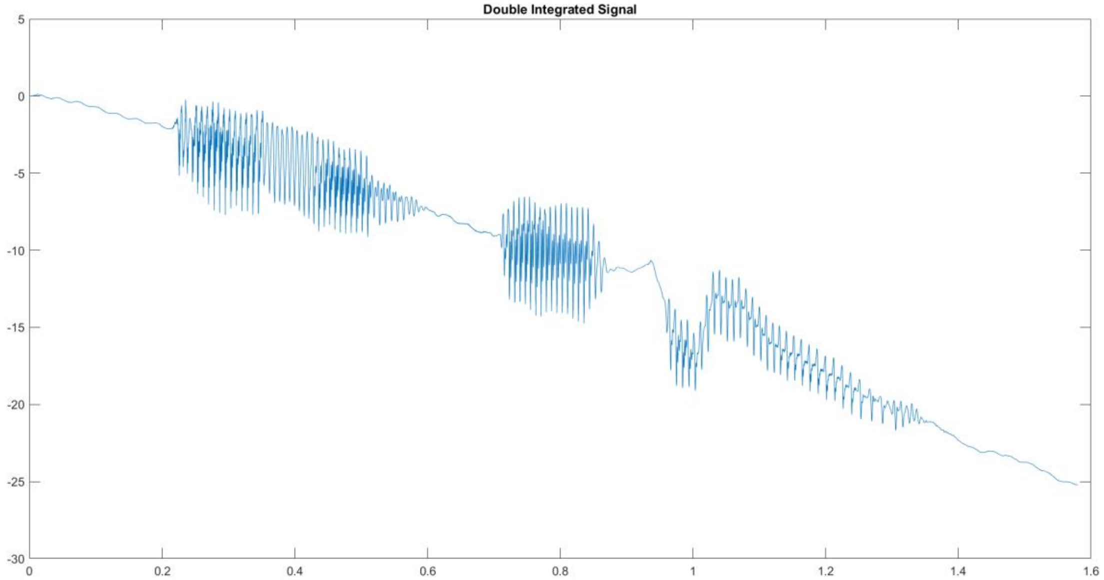
2) $y_1[n] = x[n] + 2 \cdot y_1[n-1] - y_1[n-2]$ {Double Integration}

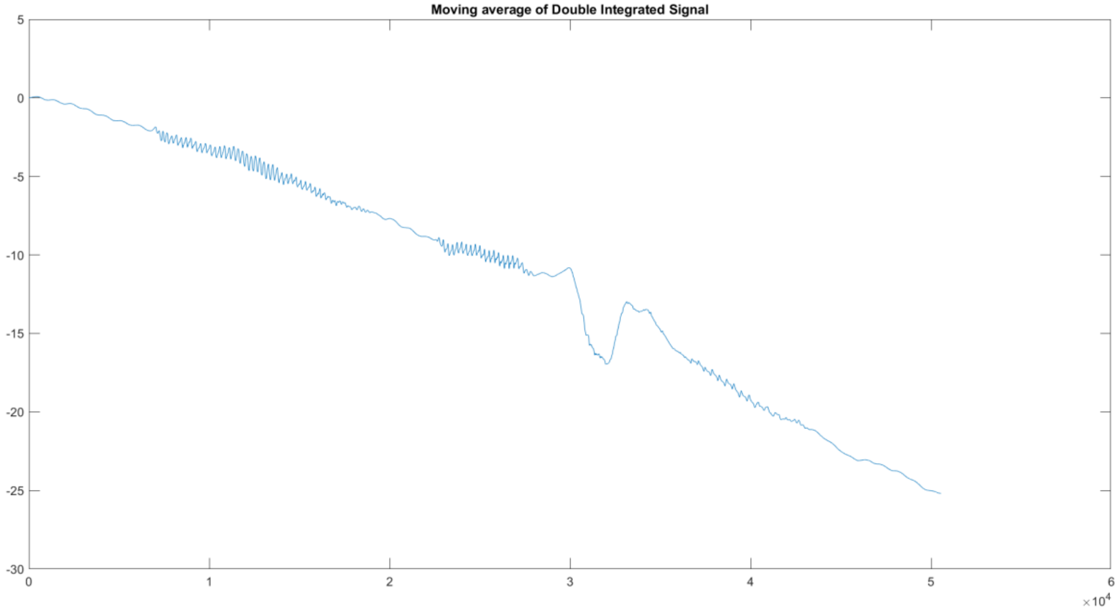
3) $y[n] = y_2[n] - \frac{1}{2N+1} \sum_{m=-N}^{N} y_2[n+m]$ {Residual of Mean Filtered signal}

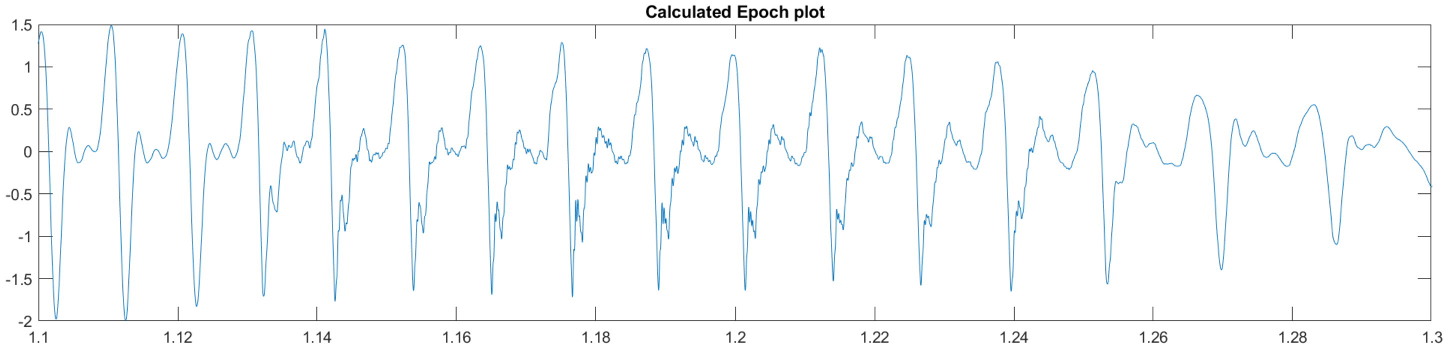
The speech signal a0030.wav file is taken from the CMU artic library and the results are as follows.

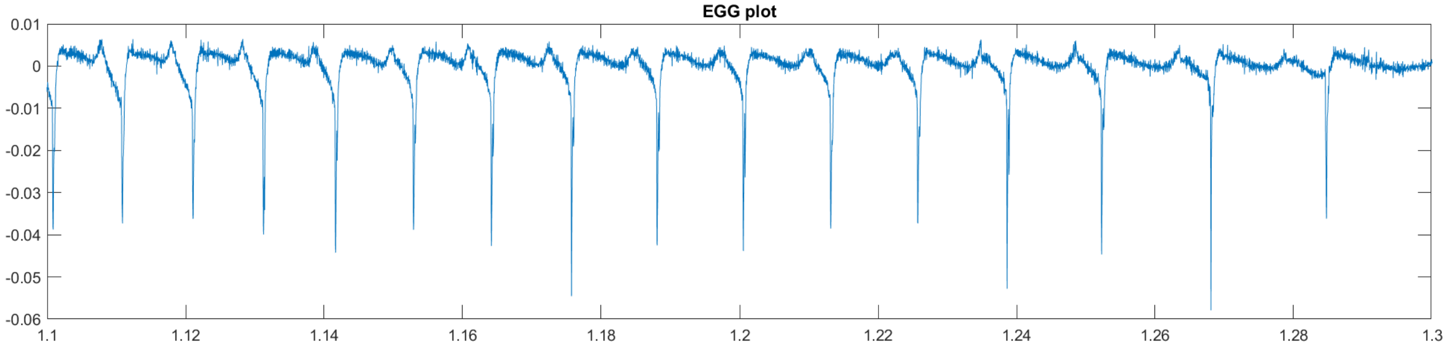



Pre-emphasised Signal


Double Integrated Signal

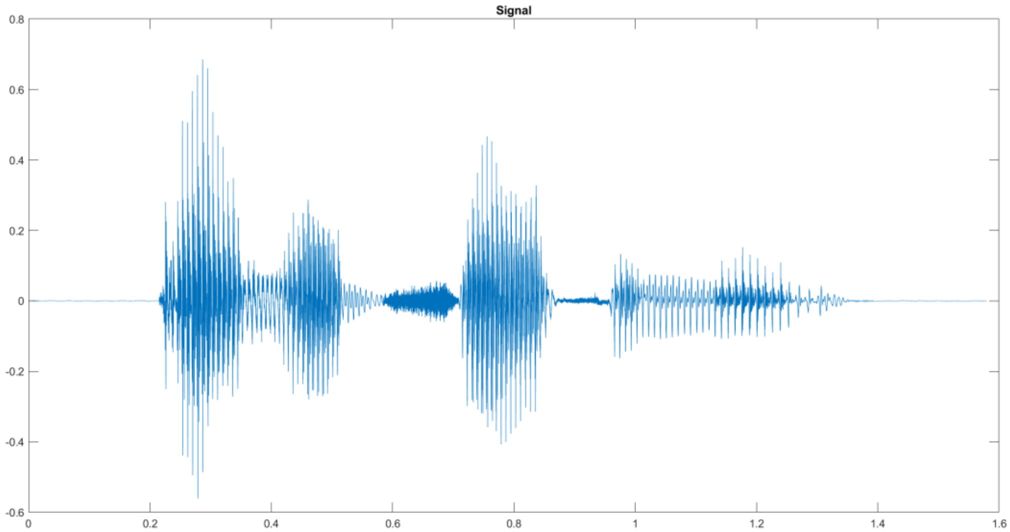

Moving average of Double Integrated Signal



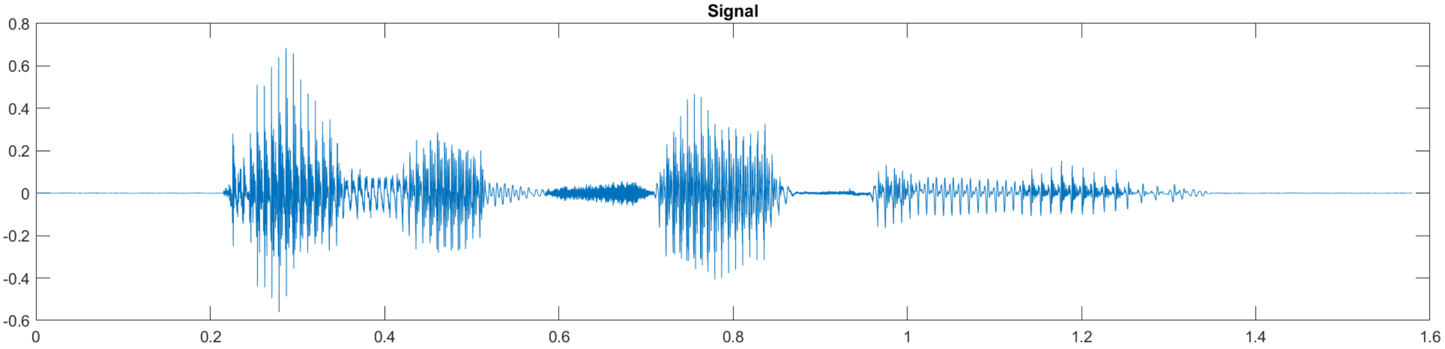


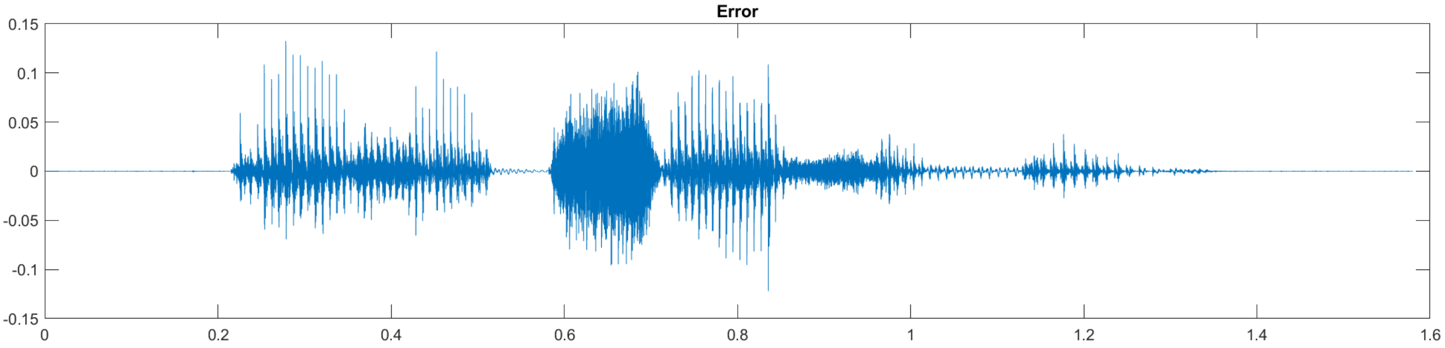
The epochs are calculated as shown above and compared to the EGG plot. The results match that of the EGG plot.

Find LP Residual for a wave file of your choice and apply MFCC on it. write your observations. Note: Computer-based Question [10 points]

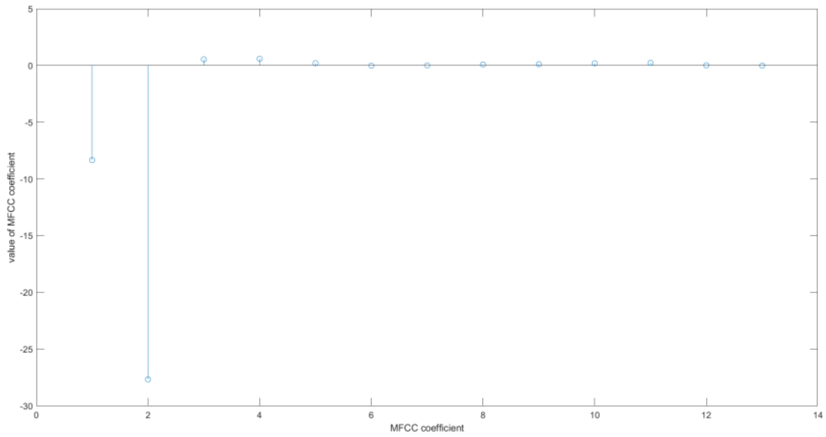


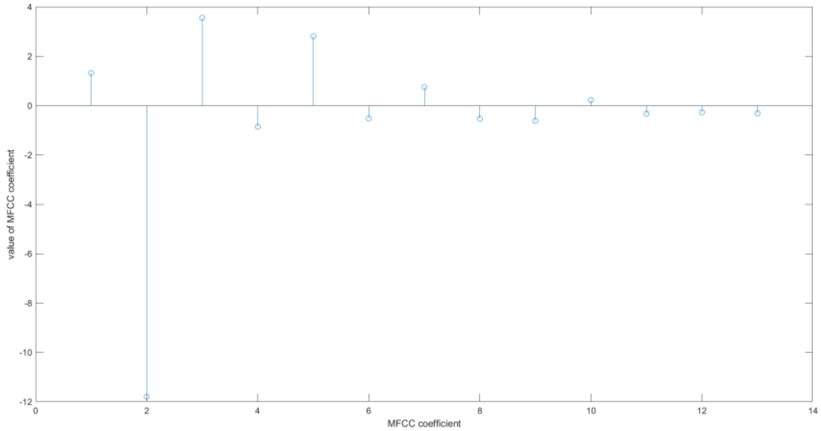
The LP residual can be found using the a_k values found from the signal and subtracting the estimated signal from the original signal. This gives the error plot which gives the excitation features of the speech signal. The plots obtained are as follows.

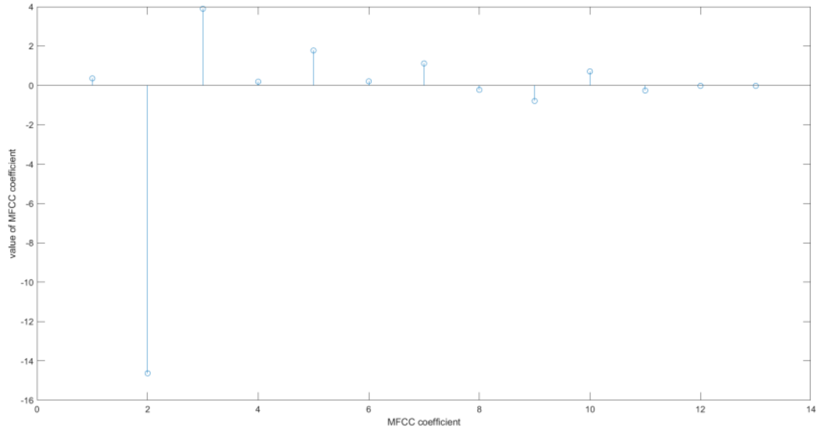


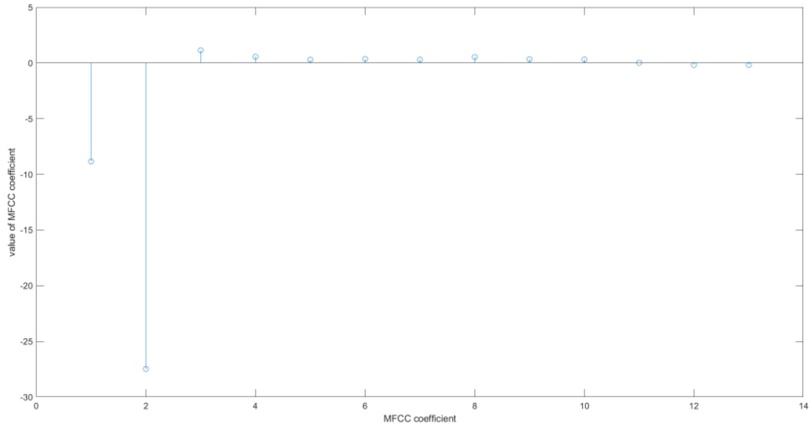


Applying MFCC for the error signal gives the following coefficients in 1st,51st,101th and 151st frame





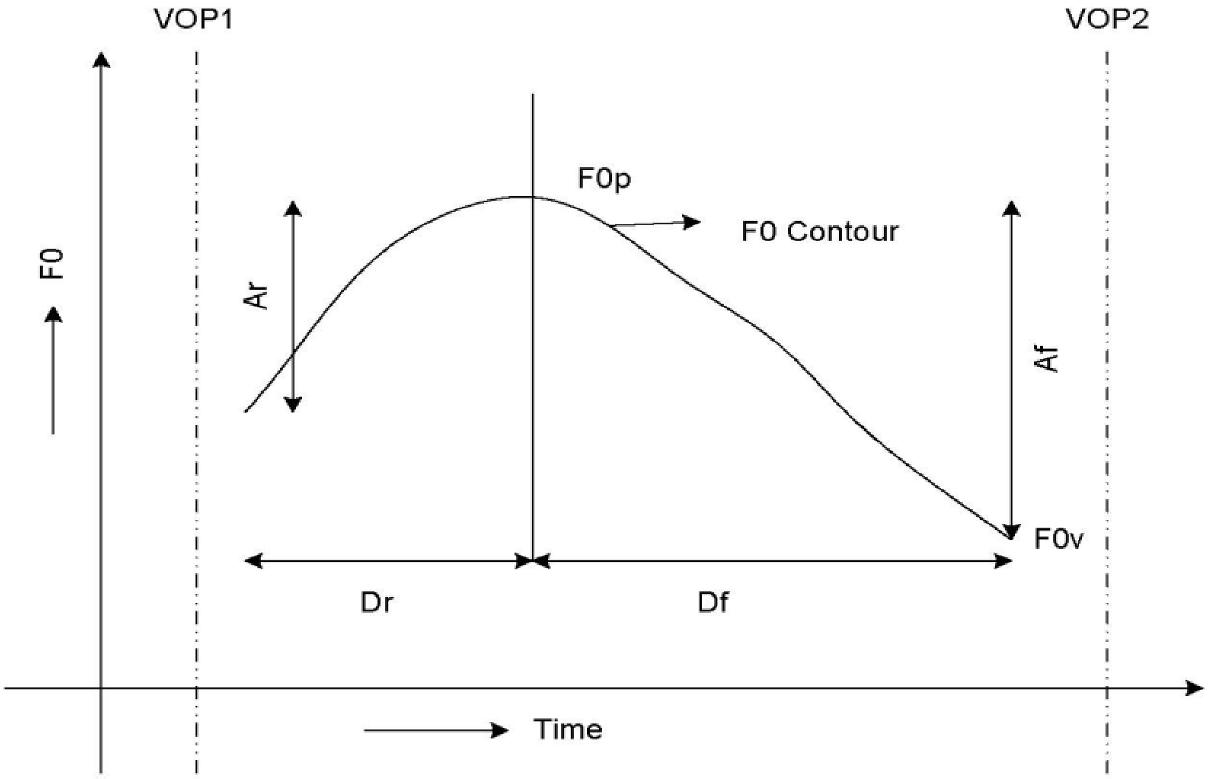




Calculate 7 prosody features for 4 wave files of the same sentence spoken by different Native speakers (Mother's tongue). Comment on variations in each feature. Note: Computer-based Question [10 points] Ex: Sentence be like "Mera Bharath Mahan". The same sentence should be recorded by 4 different native speakers.

The prosody features include

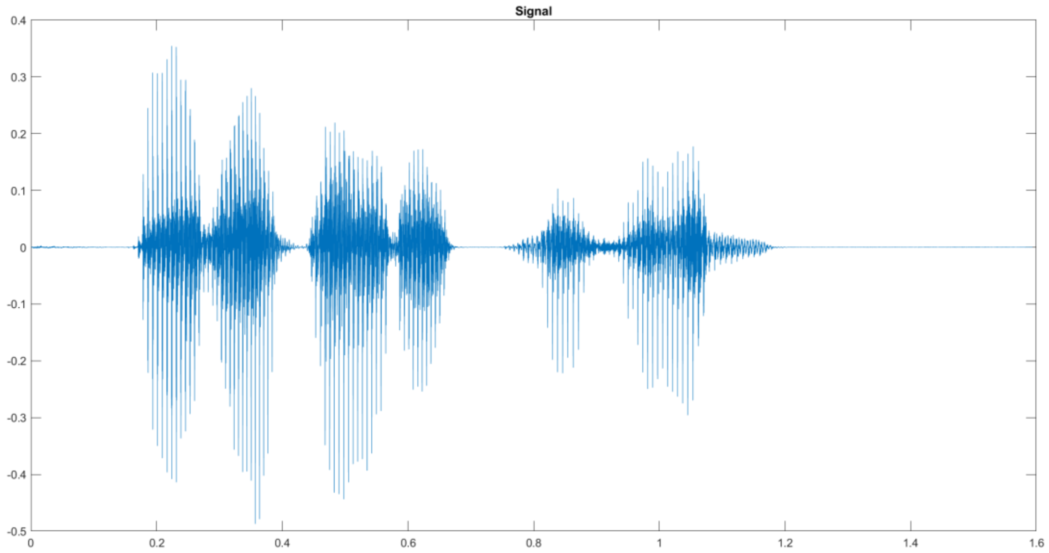
1) Change in F0: Change in the pitch of the speech signal
2) Distance of F0 peak with respect to VOP: Time between the highest point in the pitch and the VOP before or after it.
3) Amplitude Tilt: It is defined as $(A_r-A_f)/(A_r+A_f)$



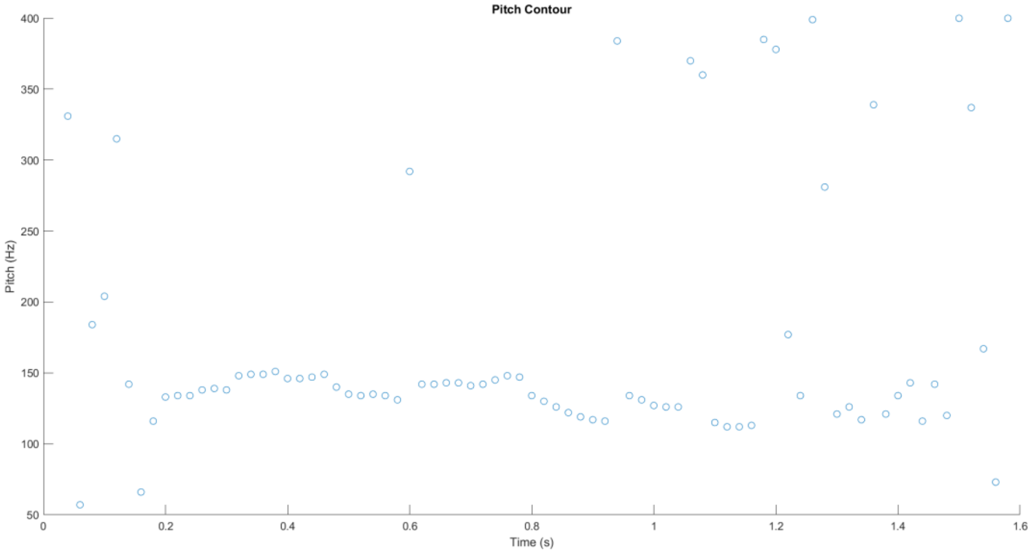
4) Duration Tilt: $(D_r-D_f)/(D_r+D_f)$
5) Distance between successive VOP: Distance between the two vowel onset points.
6) Duration of voiced region: It is calculated by removing the unvoiced and the silence part by removing the signal with less than certain energy.
7) Change in log energy in the voiced region: The log of the energy of the

The following are measured for 4 speakers with the native language being Telugu, Tamil, Hindi, and Malayalam with the sentence being used is "Mera Bharat Mahan".

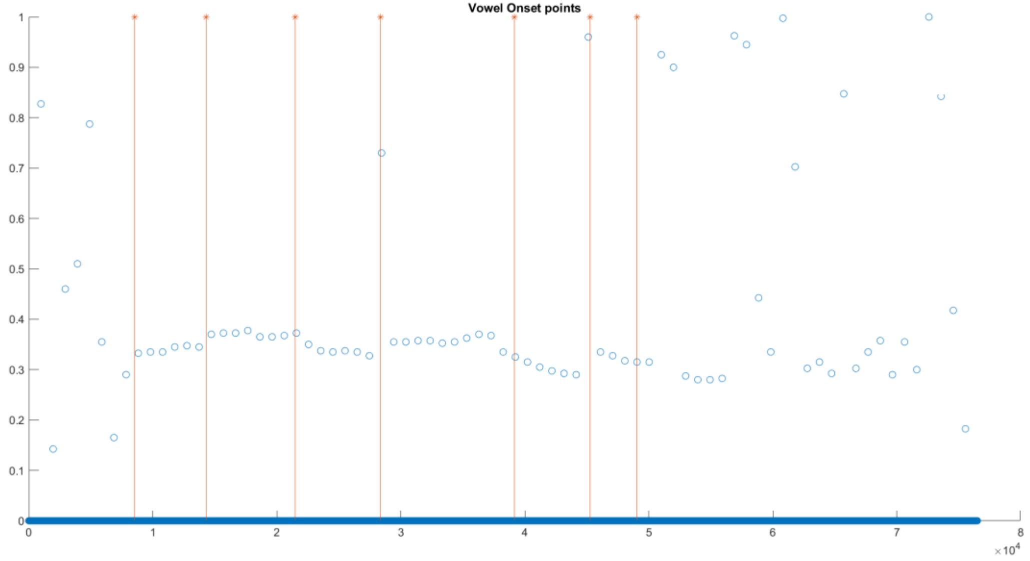
For Telugu speaker:



1) Change in F0



2) Distance of F0 peak with respect to VOP



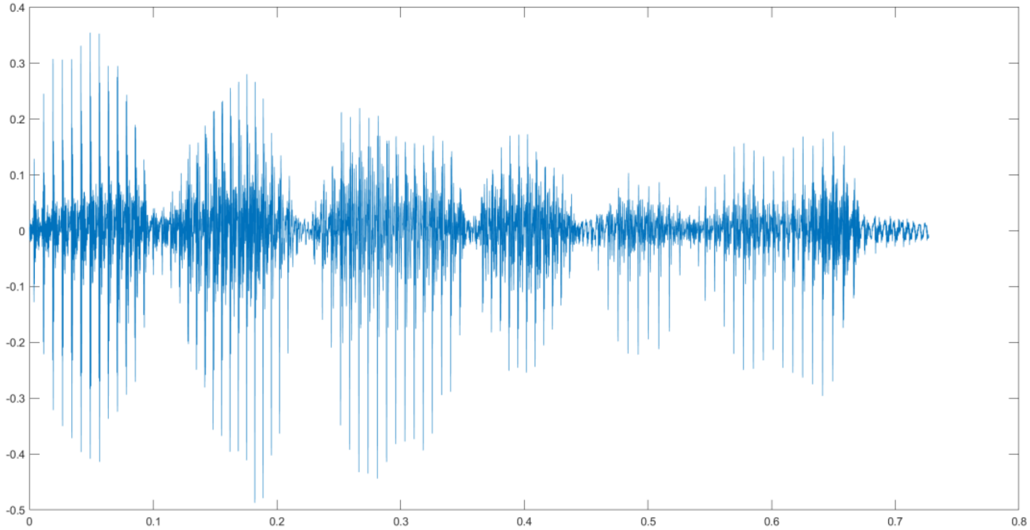
3) Amplitude Tilt
This is calculated between the first two VOP for this speech signal and the value turns out to be 0.66.

4) Duration Tilt
This is calculated between the first two VOP for this speech signal and the value turns out to be 0.13.

5) Distance between successive VOP

This is calculated between the first two VOP for this speech signal and the value turns out to be 120ms.

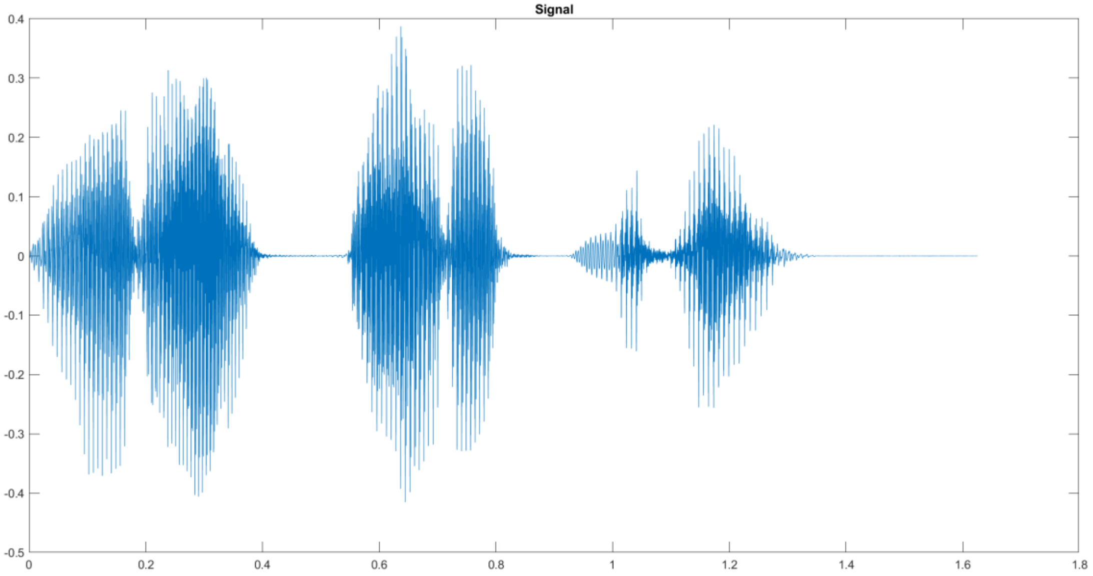
6) Duration of voiced region



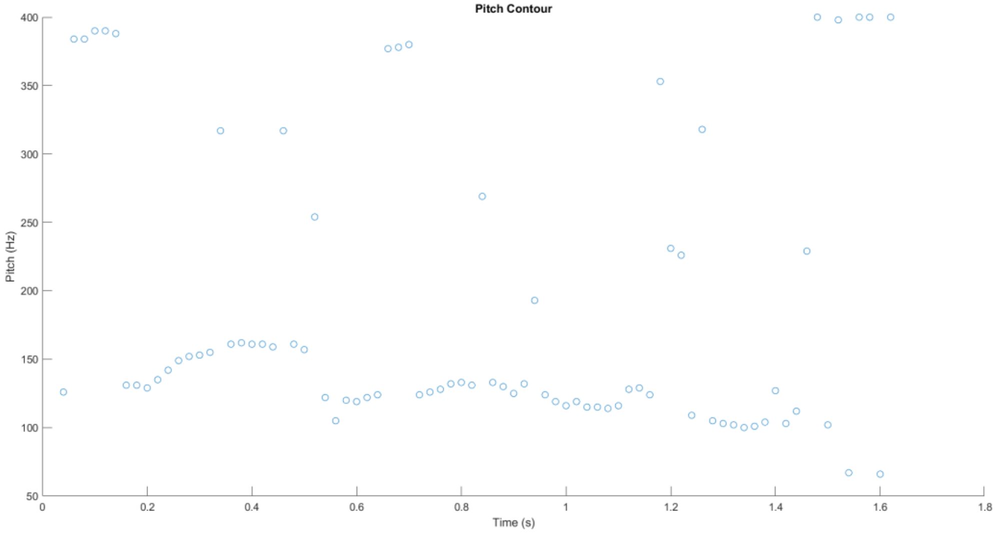
7) Change in log energy in the voiced region

```
Log Energy = 4.7994
```

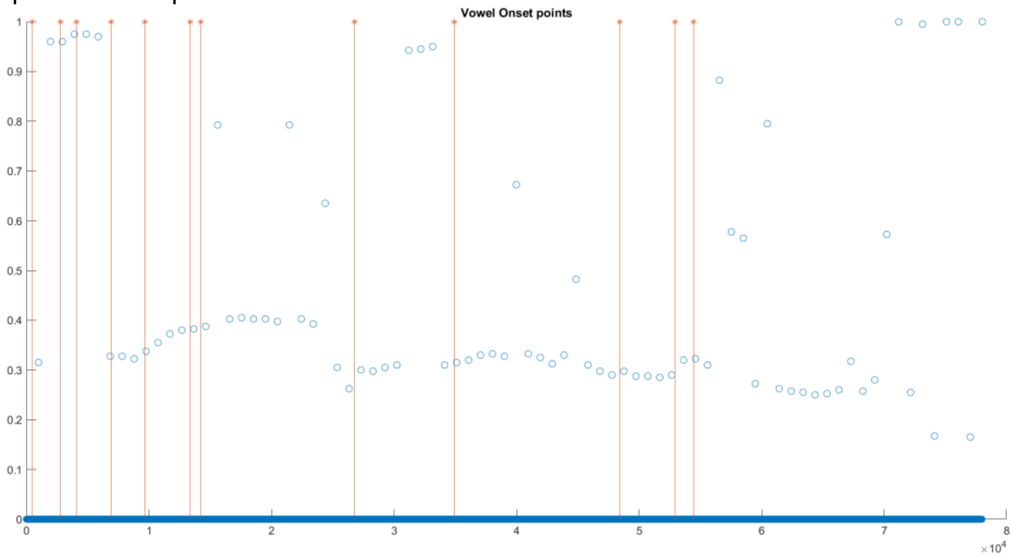
For Tamil speaker:



1) Change in F0



2) Distance of F0 peak with respect to VOP



3) Amplitude Tilt

This is calculated between the first two VOP for this speech signal and the value turns out to be 0.98.

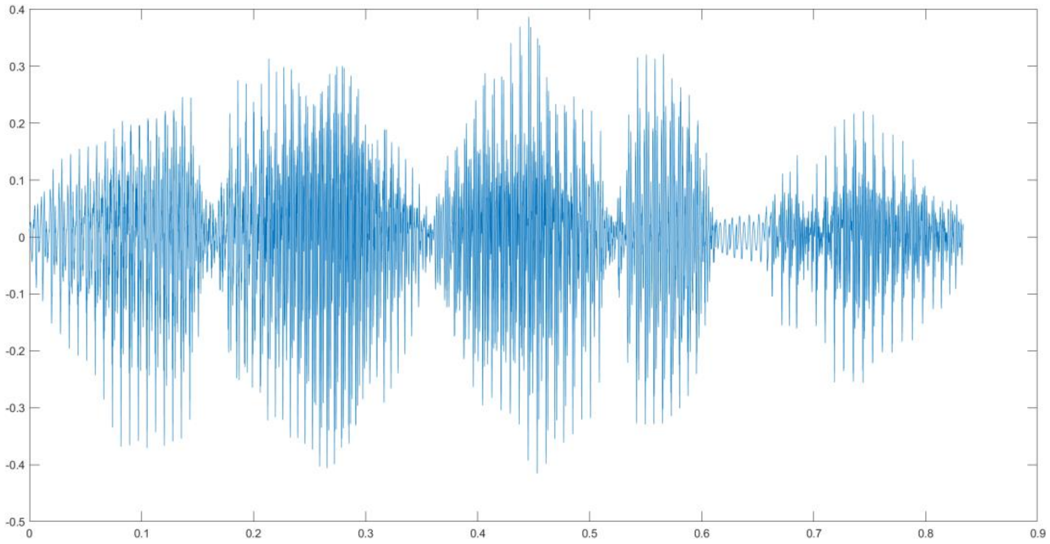
4) Duration Tilt

This is calculated between the first two VOP for this speech signal and the value turns out to be 0.286.

5) Distance between successive VOP

This is calculated between the first two VOP for this speech signal and the value turns out to be 48ms.

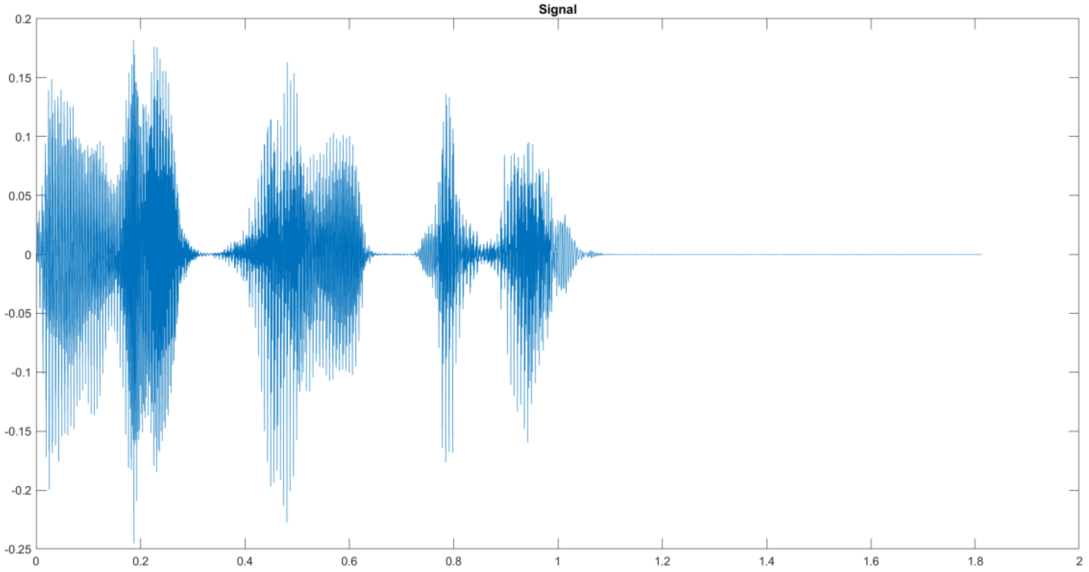
6) Duration of voiced region



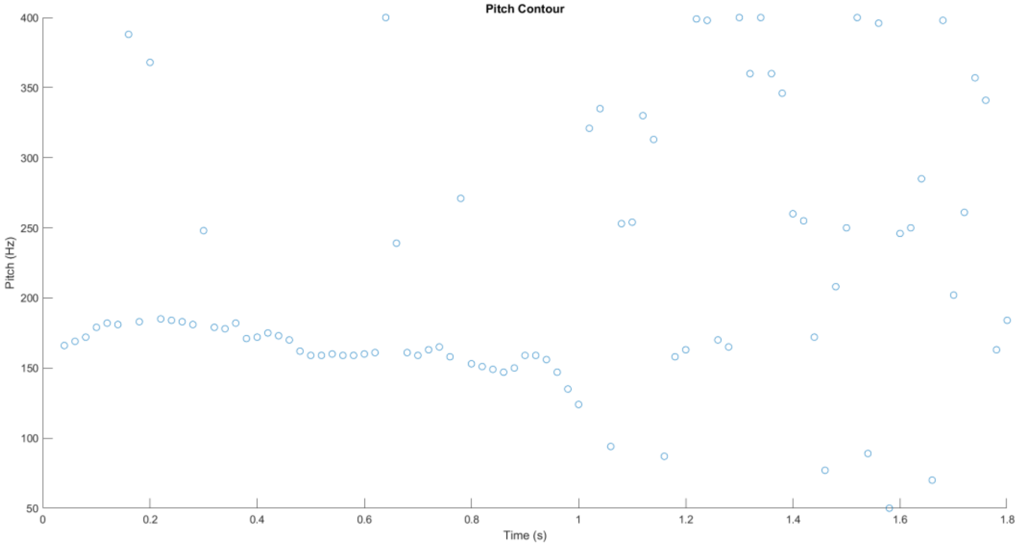
7) Change in log energy in the voiced region

```
Log Energy = 5.9970
```

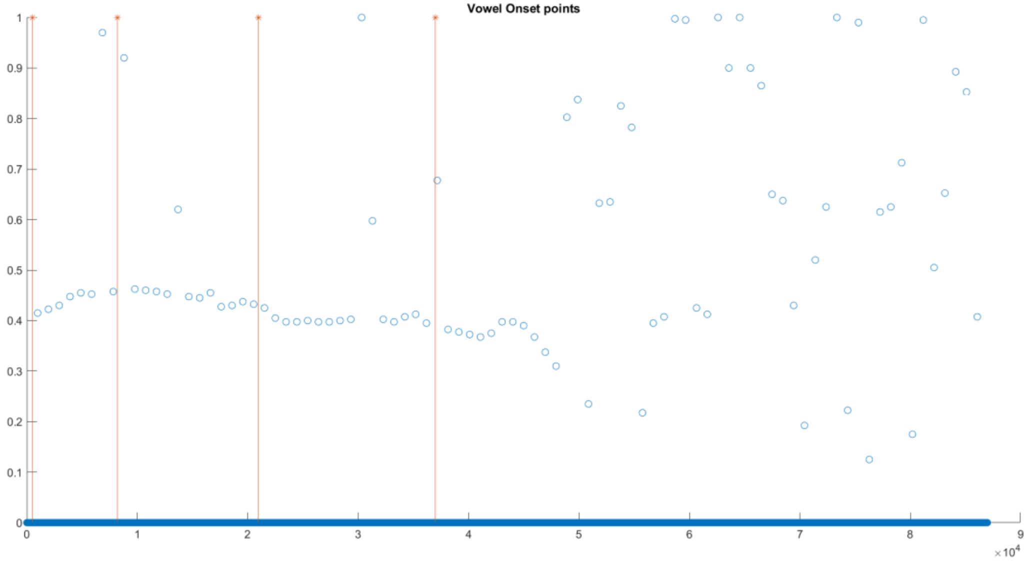
For Hindi speaker:



1) Change in F0



2) Distance of F0 peak with respect to VOP



3) Amplitude Tilt

This is calculated between the first two VOP for this speech signal and the value turns out to be 0.88.

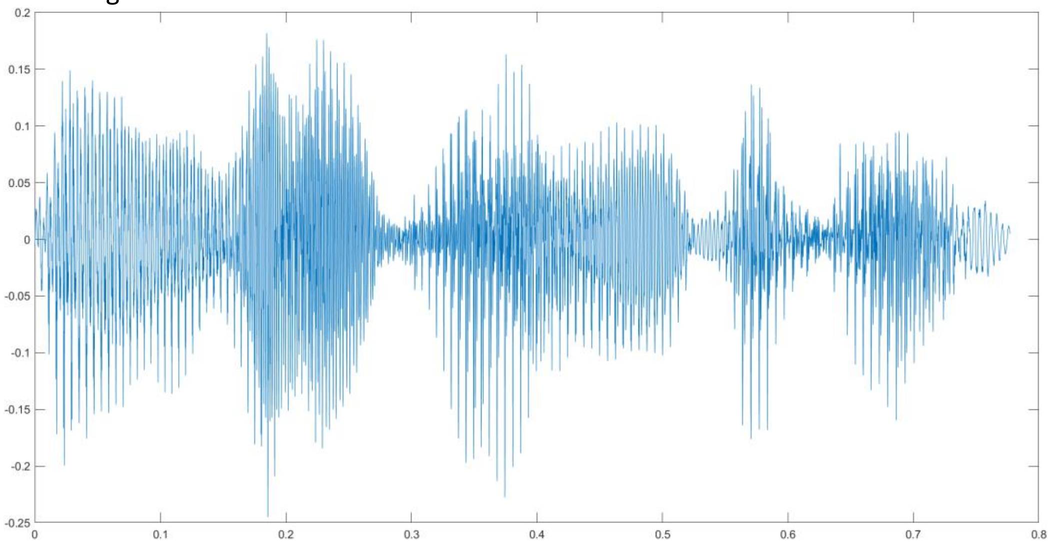
4) Duration Tilt

This is calculated between the first two VOP for this speech signal and the value turns out to be 0.14.

5) Distance between successive VOP

This is calculated between the first two VOP for this speech signal and the value turns out to be 160ms.

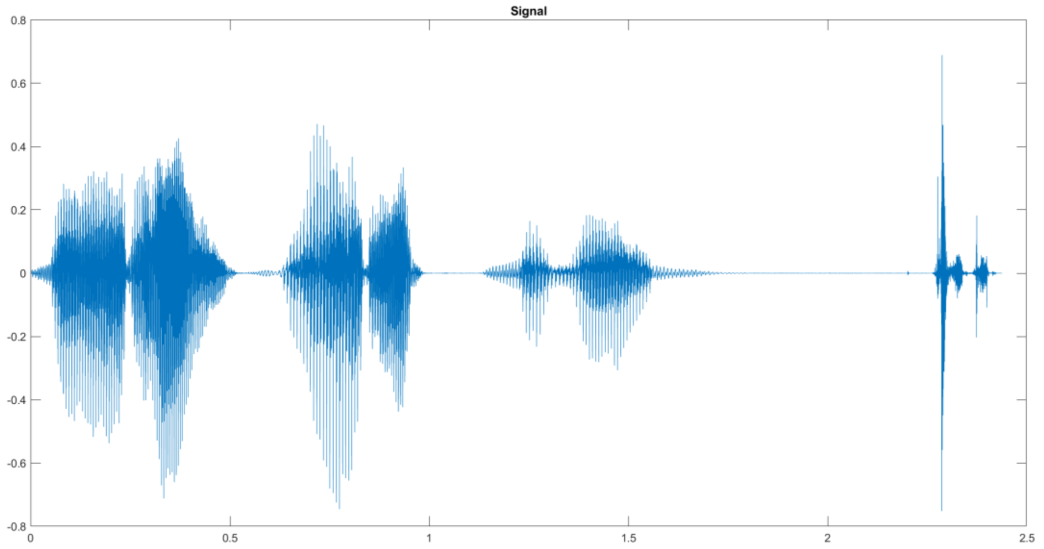
6) Duration of voiced region



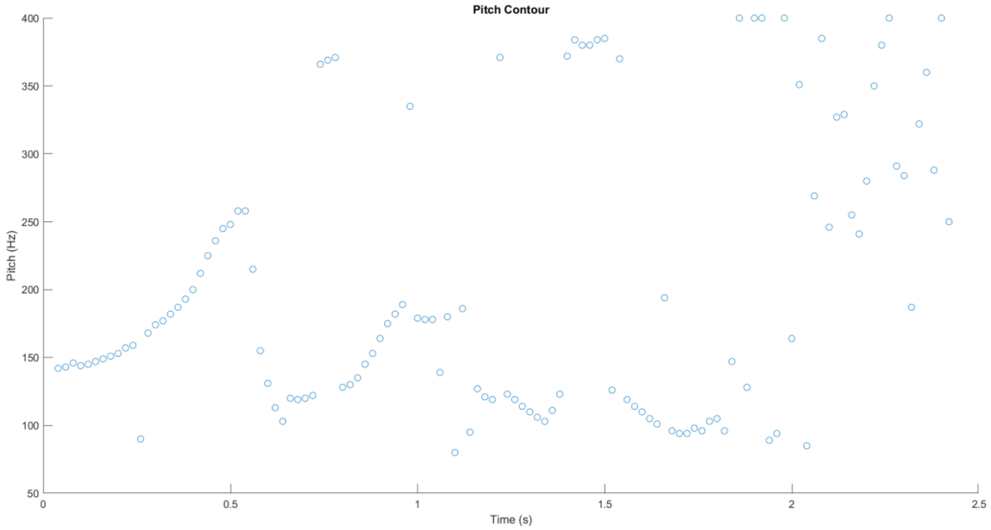
7) Change in log energy in the voiced region

```
Log Energy = 4.4810
```

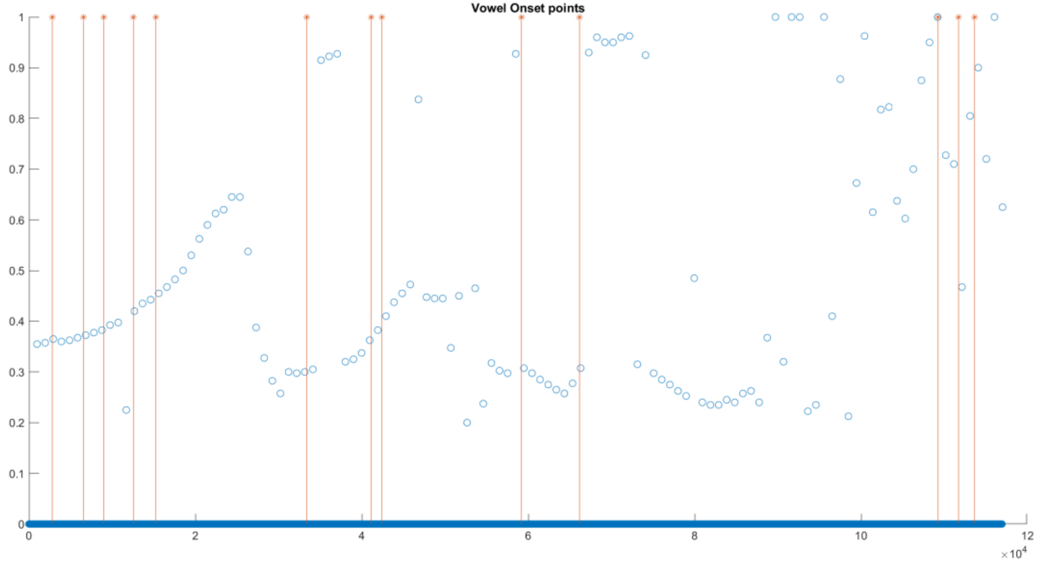
For Malayali speaker:



1) Change in F0



2) Distance of F0 peak with respect to VOP



3) Amplitude Tilt

This is calculated between the first two VOP for this speech signal and the value turns out to be 0.4285.

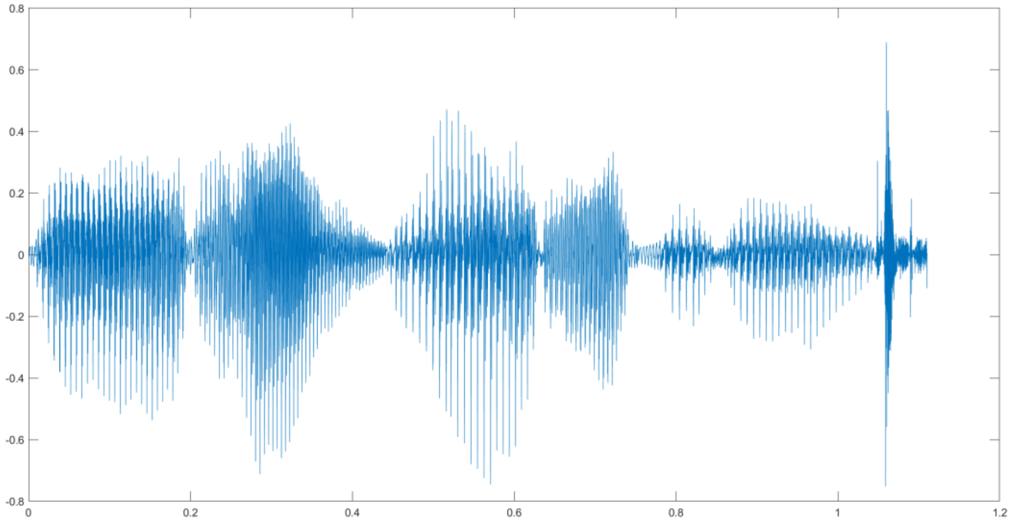
4) Duration Tilt

This is calculated between the first two VOP for this speech signal and the value turns out to be 0.42.

5) Distance between successive VOP

This is calculated between the first two VOP for this speech signal and the value turns out to be 77ms.

6) Duration of voiced region



7) Change in log energy in the voiced region

```
Log Energy = 6.5610
```

As we can see from the above plots and the results, the following table summarizes the variation in the prosody features

Native Language	Change in F0(Average pitch of each speaker is as shown)	Distance of F0 peak with respect to VOP	Amplitude Tilt	Duration Tilt	Distance between successive VOP	Duration of voiced region	Change in log energy in the voiced region
Telugu	173	70ms	0.66	0.13	120ms	0.75 s	4.7994
Tamil	186.5875	69ms	0.98	0.286	48ms	0.85 s	5.997
Hindi	215.8876	33ms	0.88	0.14	160ms	0.79 s	4.4810
Malayalam	203.325	210ms	0.42	0.4285	77ms	1.1s	6.5610