DSP PROJECT

VOWEL ONSET POINT DETECTION USING MODULATION SPECTRUM ENERGY

Mentor: Dr Anil Kumar Vuppala & Krishna Gurugubelli sir

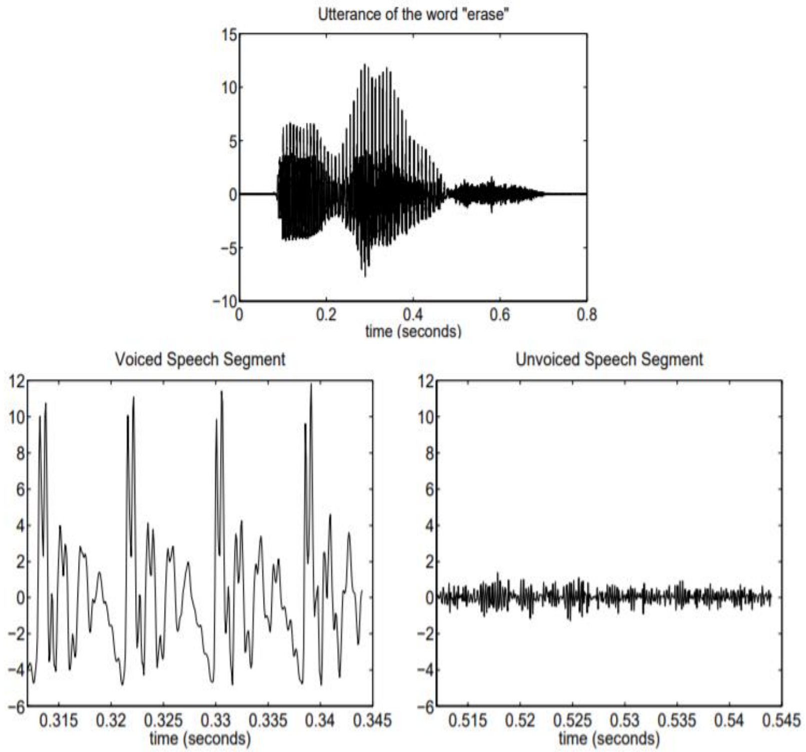
Team Members :
Salay Jain (20171078)
Aradhya Tongia (20171049)

Project Description:

- ❖Vowel Onset Point is the instant at which the beginning of a vowel takes place during speech production.
- ❖Vowels are produced by keeping the vocal tract in an open position with minimum obstruction along the length and using glottal vibration as the excitation.
- ❖A speech signal is composed of voiced signal, unvoiced signal and noise. Vowels along with few consonants a considered as voiced while others as unvoiced.
- ❖Voiced speech is produced by vibrations in vocal cords thus can have high energy compared to unvoiced speech which are low energy.
- ❖Therefore voiced signals tend to be louder like the vowels /a/, /e/, /i/, /u/, /o/. Unvoiced signals, on the other hand, tend to be more abrupt like the stop consonants /p/, /t/, /k/.
- ❖So whenever vowel is produced there must be significant change in energy thus we now know to identify we will analyze the energy of the speech signal and process it to find VOP.
- ❖There are significant changes occurring in the energies of excitation source, spectral peaks, and modulation spectrum at the point when a vowel is pronounced in speech.
- ❖So our project is to detect Vowel onset point using modulation spectrum energy.

Significance of using modulation spectrum energy:

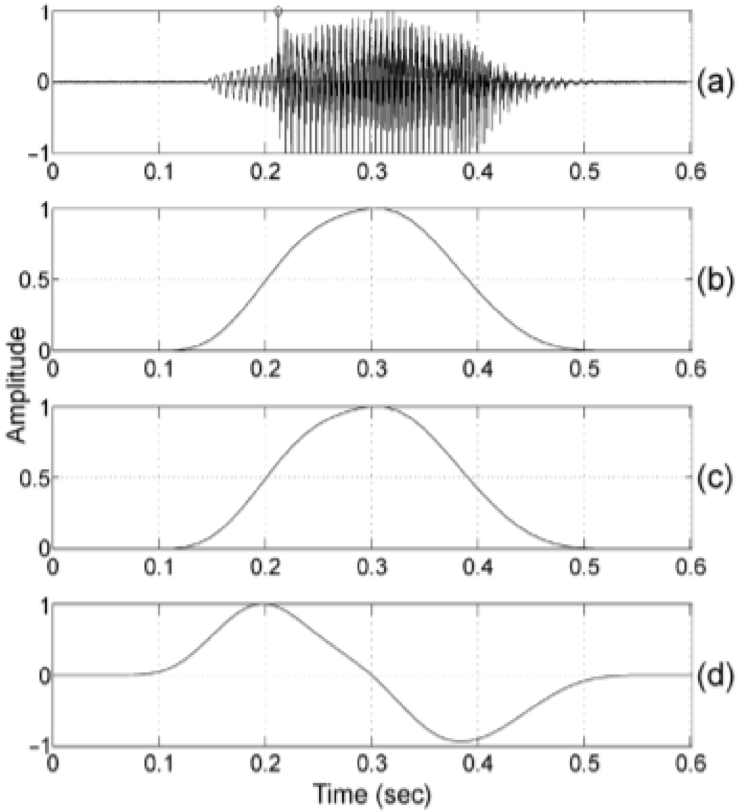
- ❖ Several studies have been conducted earlier to demonstrate the significance of modulation spectrum in speech processing
- ❖ Modulation components refer to the slowly varying temporal envelope components in speech.
- ❖ Temporal envelope of speech is dominated of low frequency components by several Hertz.
- ❖ It is also demonstrated that the concentration of energy in the modulation spectrum corresponds to the syllable nuclei that is vowel.
- ❖ Thus, the onset of vowel may be manifested as a significant change in the modulation spectrum energy level in the 4–16 Hz band.



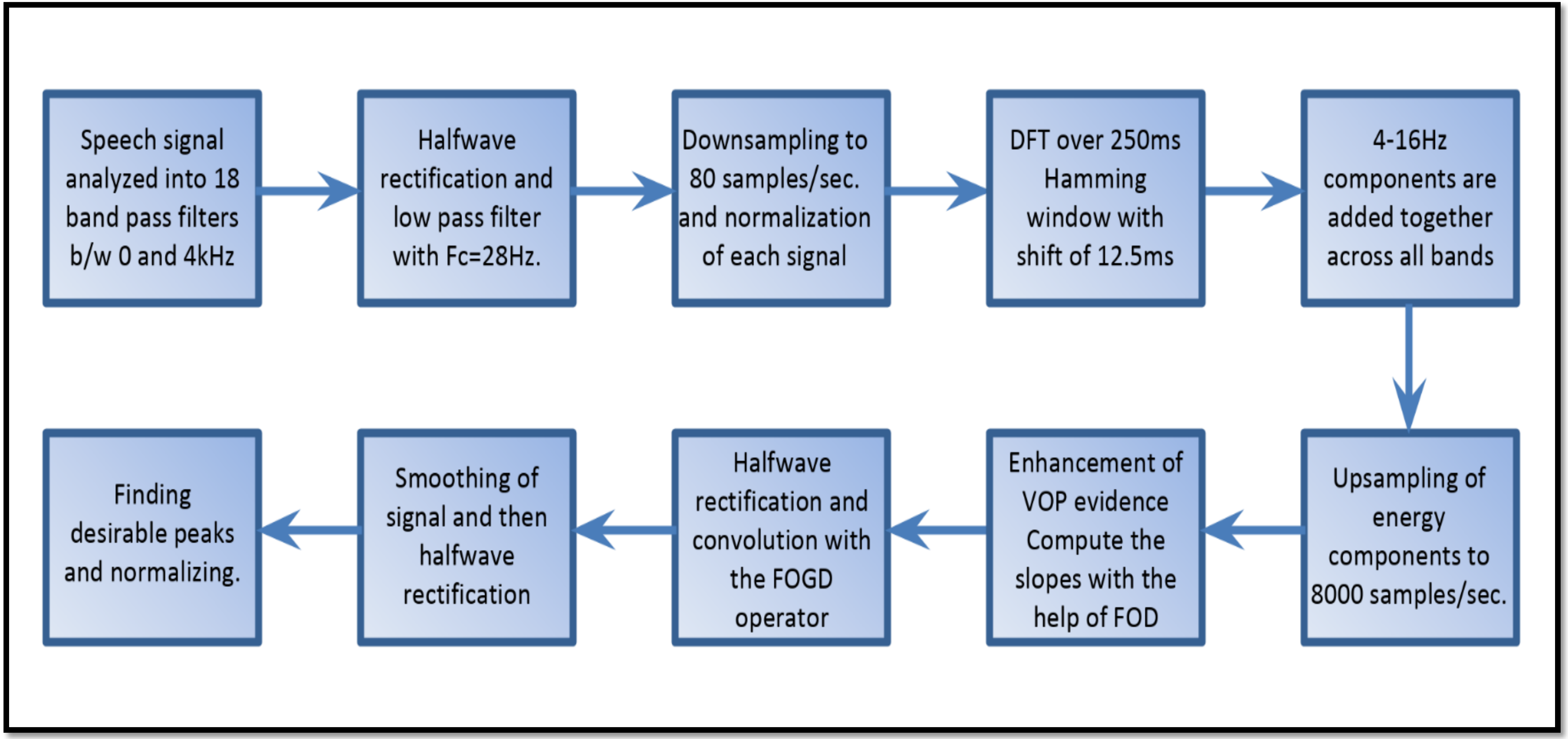
Understanding the problem:

Lets consider this speech signal [Figure (a)] for word ‘na’:

- ❖ As we know ‘n’ being unvoiced has low energy while ‘a’ has high energy.
- ❖ Figure (b) shows the modulation spectrum energy.
- ❖ Figure (c) shows the enhanced modulation energy spectrum which we get by smoothening the modulation spectrum energy.
- ❖ Figure (d) show the VOP evidence plot which we get using FOD (first order differential operator) and FOGD (first order gaussian differential operator).
- ❖ Thus the vowel onset point is detected for this speech signal.



Block diagram:



Mathematics involved:

For finding the energy distribution for the 18 bands we use the equation:

$$m(i) = \sum_{p=1}^{18} \left[\sum_{k=k_1}^{k=k_2} \left| \hat{X}_p(k,i) \right|^2 \right]$$

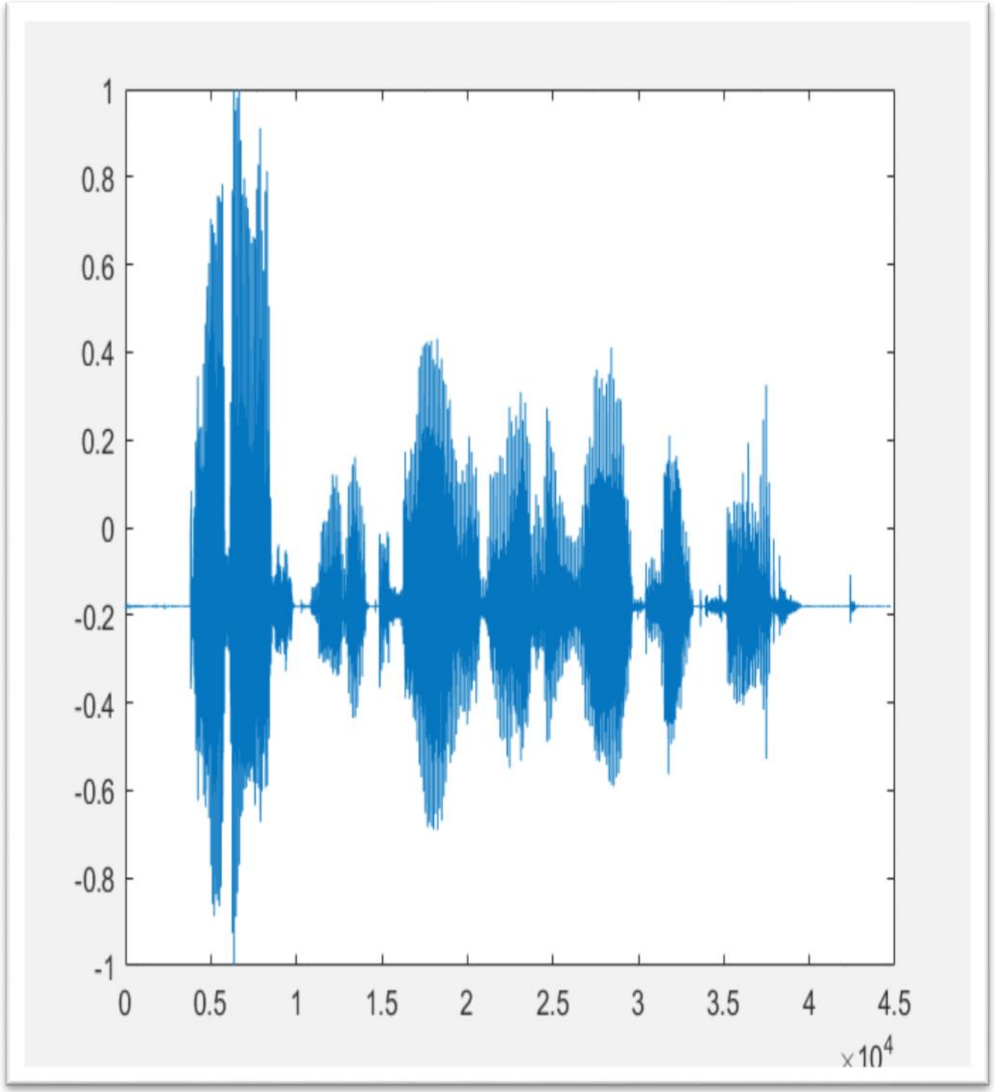
where:

$$\hat{X}_p(k) = \sum_{n=0}^{N-1} \hat{x}_p(n) w(n) e^{-\frac{j2\pi nk}{N}}; \quad p = 1, 2, \ldots, 18.$$

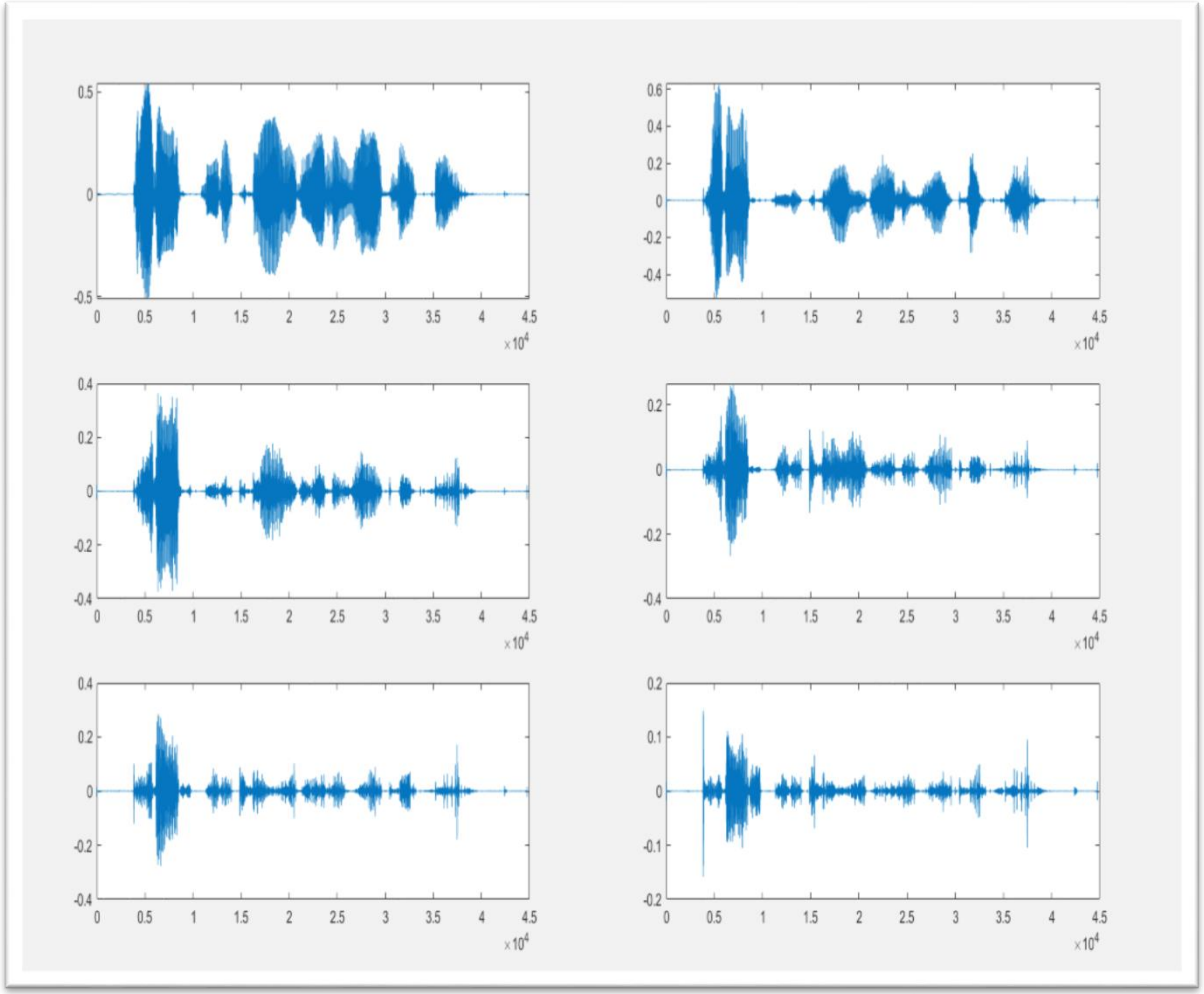
Since we got the energy distribution now, we need to further enhance our energy spectrum to remove close peaks and smoothen the signal.

The change in VOP is further enhanced by computing its slope using FOD (first order differential).
These values are then convolved with FOGD operator (first order gaussian differentiator) and the output is the VOP evidence.

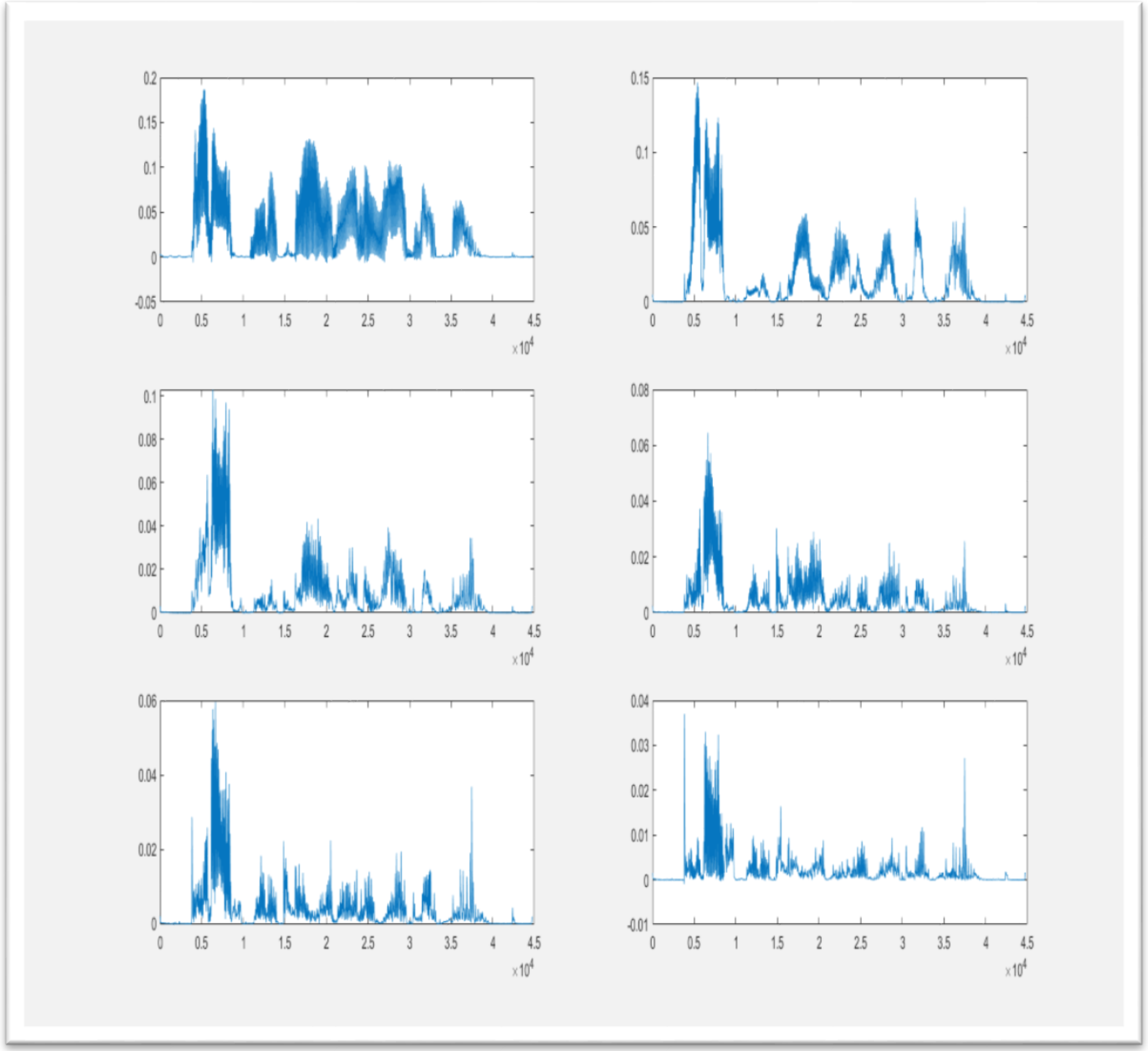
Thus our VOP evidence plot show us the various vowel onset points in our speech signal.



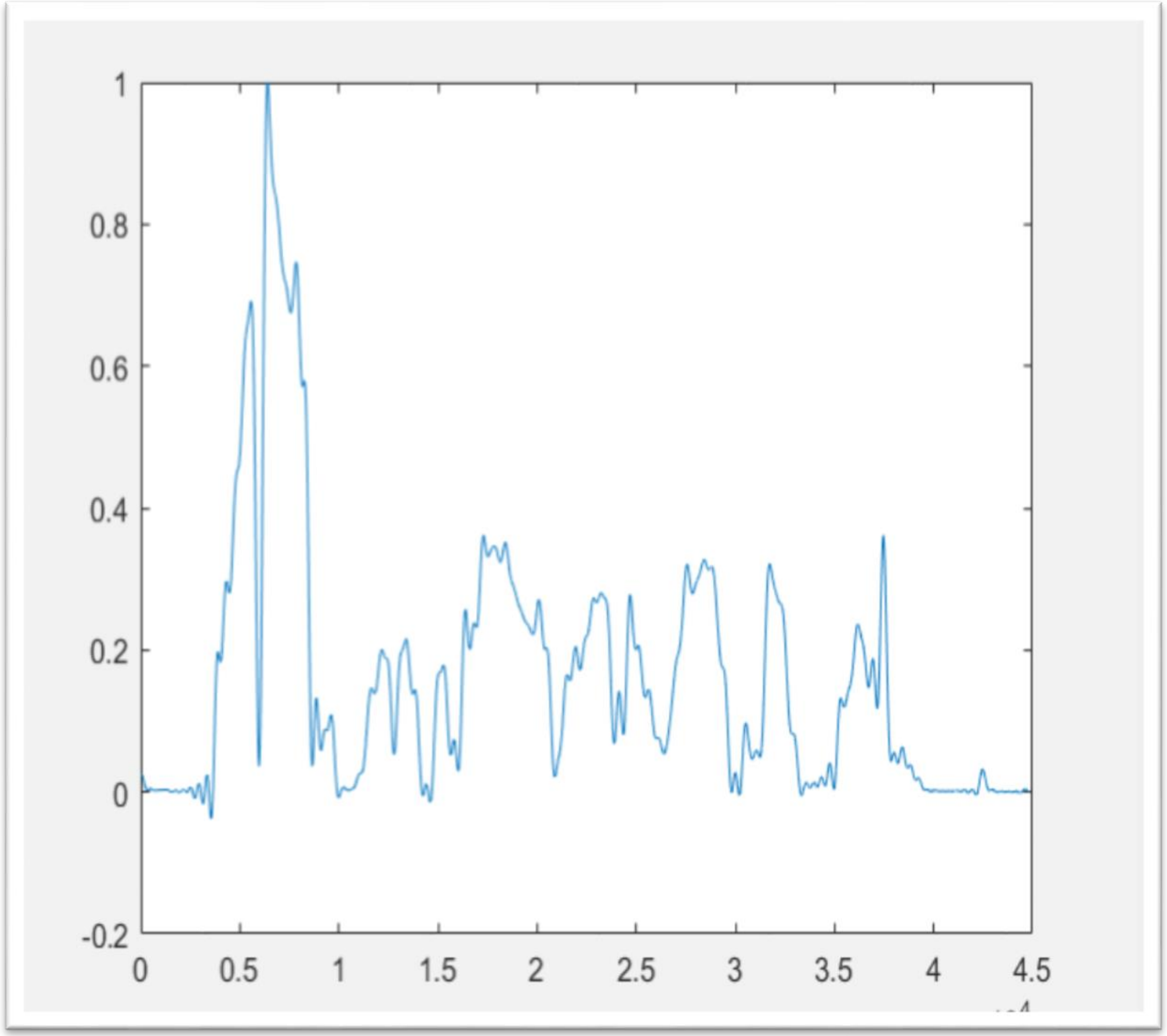
Speech signal
“Dont_ask_me_to_carry_an_oily_rag_like_that”



Speech signal analysed to 18 bandpass filters between 0-4KHz.



Half-wave rectification and low pass filter of 28Hz for all the 18 bands.



All the 4-16 Hz energy components are added to get the final energy band.

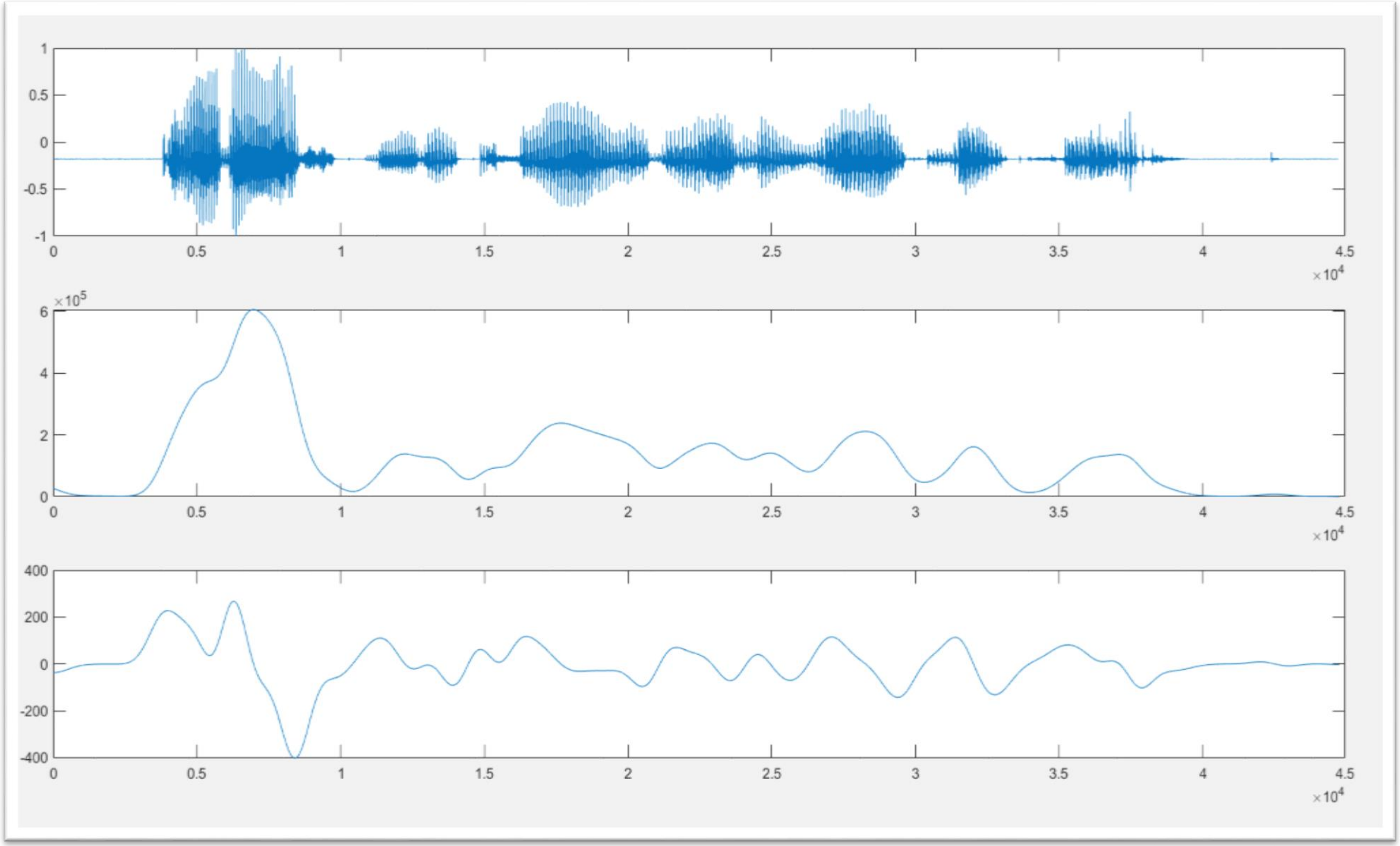


Figure 1:
Speech signal to identify VOP's.

Figure 2:
Energy components of 4-16 Hz are added and smoothen for all the 18 bands.

Figure 3:
First order Differential of the energy band after smoothening.

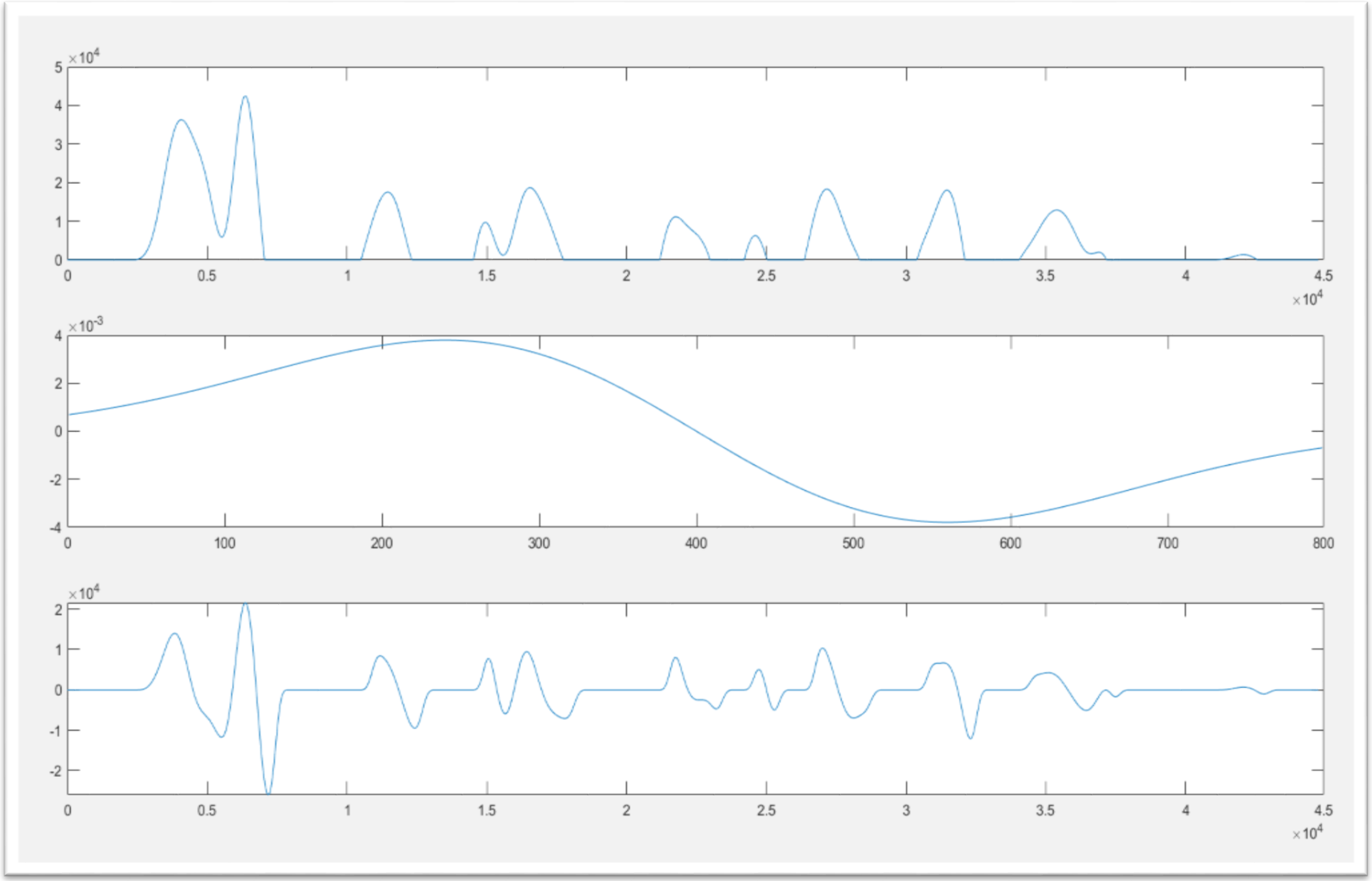


Figure 1:
Sum of peaks for 10 mS window period is taken and rectified.

Figure 2:
FOGD operator with which we will convolve the above signal to get accurate VOP's.

Figure 3:
After applying first order gaussian differential operator we get enhanced VOP.

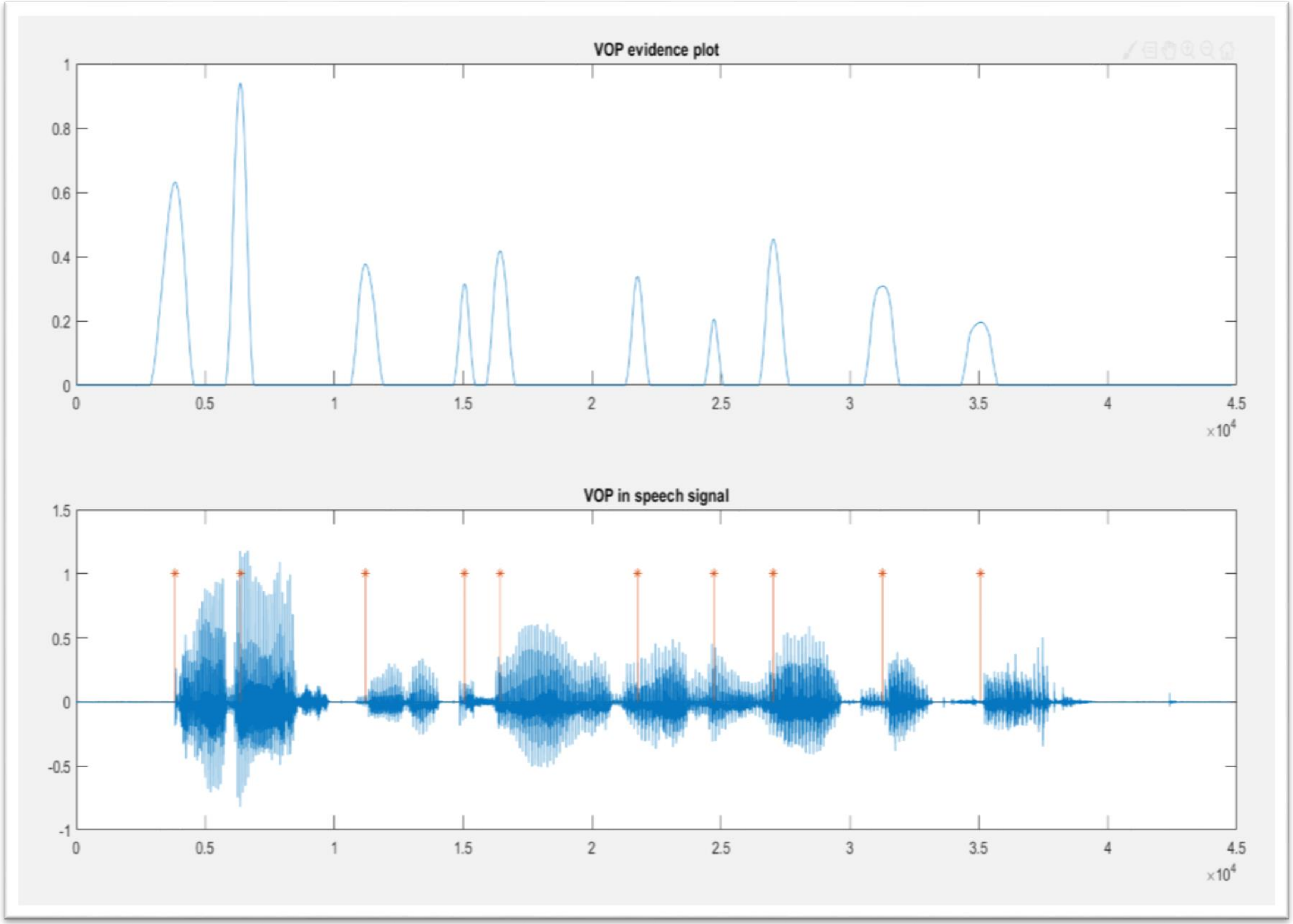


Figure 1:
VOP evidence plot after half wave rectification and normalizing.

Figure 2:
Peaks in VOP evidence plot gives us vowel onset point in our speech signal. Thus all orange ‘*’ denotes the VOP in our signal.

Applications

- Used for consonant vowel (CV) transitions.
- Detection of end-points of speech utterance.
- Is used as anchor points to extract features for speech recognition.

Drawbacks:

- Noise has a significant effect on our result. If large amplitude noise is added to our signal than it may be detected in our modulation spectrum and give us wrong VOP’s.
- While using certain filters in MATLAB they create delays which may lead to shifting of our VOP’s.

Reference:

1. Vowel Onset Point Detection Using Source, Spectral Peaks, and Modulation Spectrum Energies by S. R. Mahadeva Prasanna, Member, IEEE, B. V. Sandeep Reddy, and P. Krishnamoorthy, Student Member, IEEE.
2. Vowel Onset Point Detection for Low Bit Rate Coded Speech by Anil Kumar Vuppala, Student Member, IEEE, Jainath Yadav, Saswat Chakrabarti, Member, IEEE, and K. Sreenivasa Rao, Member, IEEE