DSP Project Report

Voiced Unvoiced Speech Detection using EMD (Empirical Mode Decomposition)

Shivani Chepuri – 2018122004
Gowri Lekshmy – 20171053

Aim:

To achieve a robust threshold and label V-UV parts with respect to it.
To compare and understand this method with the others.

Why this Method?

When we use the signal's mean of energy contour as a threshold (voiced have higher energy), sometimes the unvoiced frames which are unusually of high energy may be detected as voiced. Similar is the case we use variance/correlation (correlation is more and variance is less for voiced due to harmonic/periodic nature)
Therefore, we decompose the signal into harmonics to get IMFs about fundamental frequencies and apply these above methods, which in turn gives a robust threshold.

Implementation:

- Apply EMD on a speech signal to obtain IMFs.
- Select IMFs [3:5 (or) 4:6] around fundamental frequency of speech.
- Pass these IMFs into the "energy_cont_ol" function for further processing.
- On each IMF, Use "buffer" command and obtain frames of 20 ms each.
- Then, get energy and variance contours by operation of those frames and smooth the waves using "filtfilt" command and hamming window.
- Apply auto-correlation on IMFs and get their fractional energies.
- Use these Fractional Energies to get a mean fractional energy which is used as the decision_factor/threshold.
- This decision factor is applied on the energy/variance contours to get V-UV parts.

HOW TO RUN?

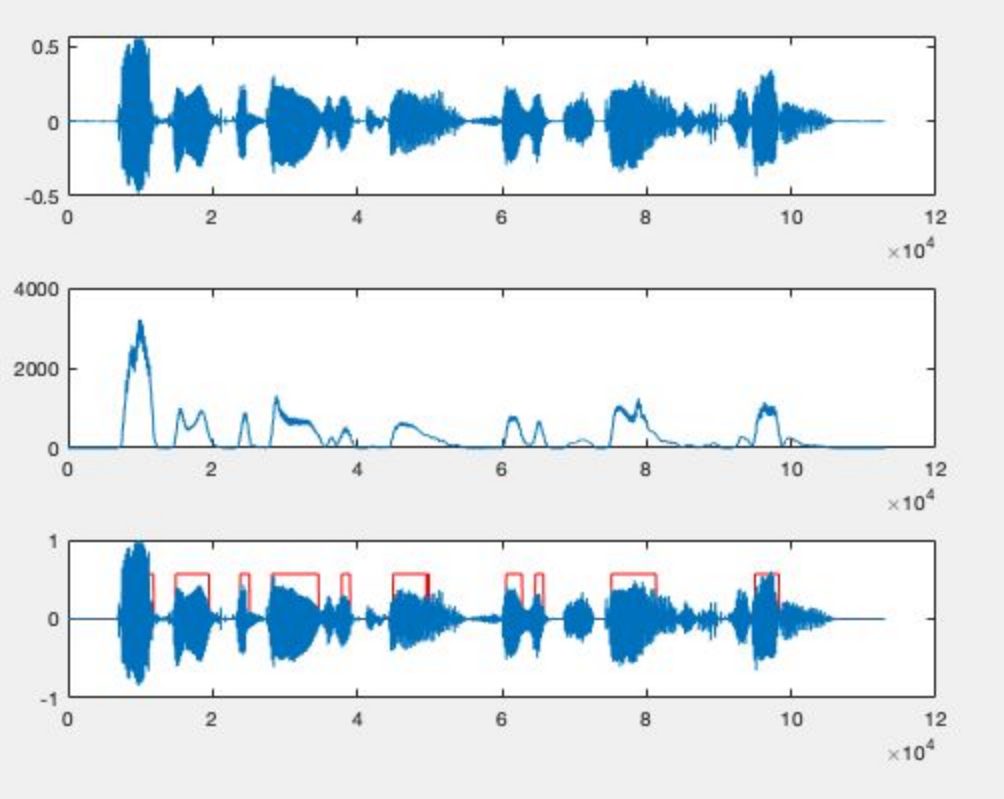
- Run "Q1.m" (It calls the "energy_cont_ol.m" whenever necessary).
- Keep the files "test.wav" and "a0002.wav" in the same directory and change the test case name in the code respectively.

.

Results:

Test case 1:

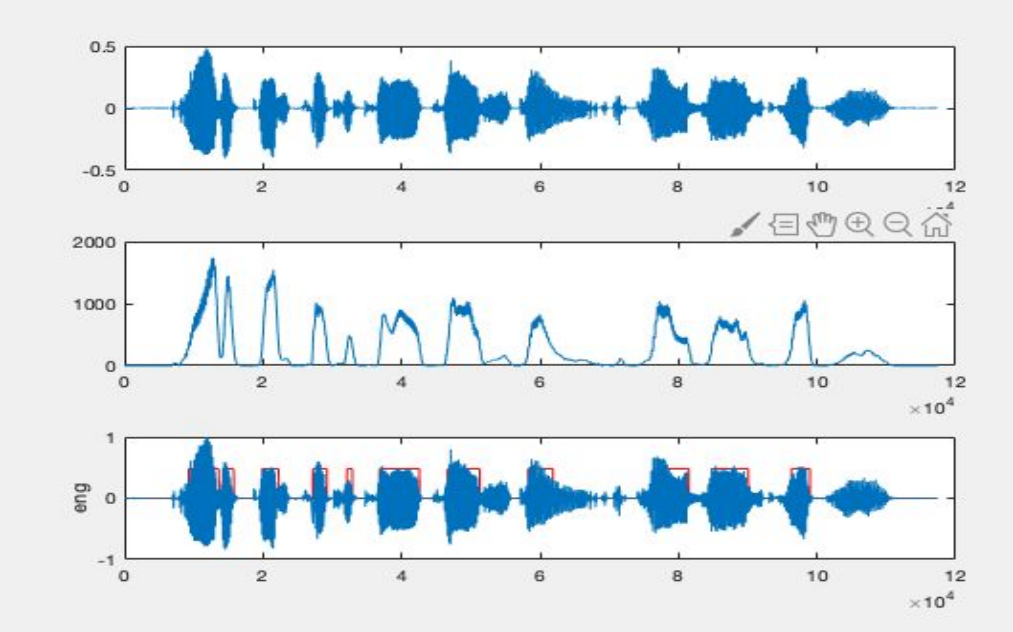
Duration: 4 sec ; Frame size: 20ms

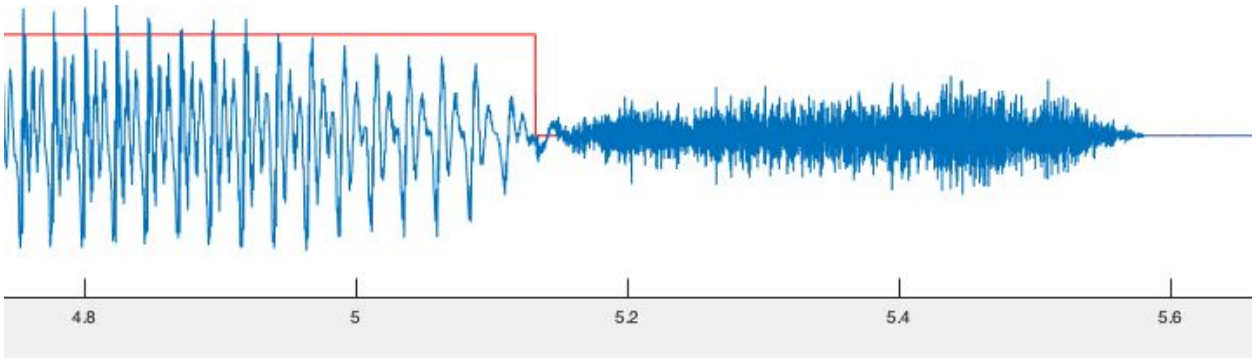


Test case 2:

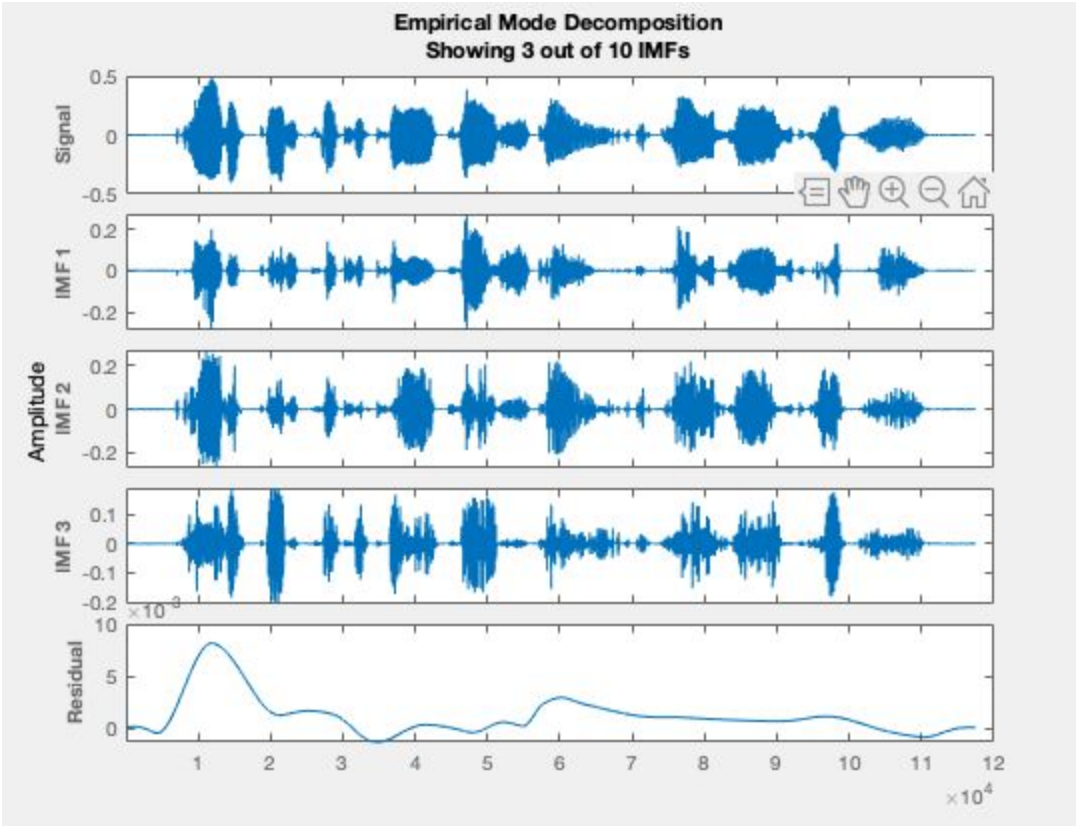
Duration: 3.6 sec ; Frame size: 20ms

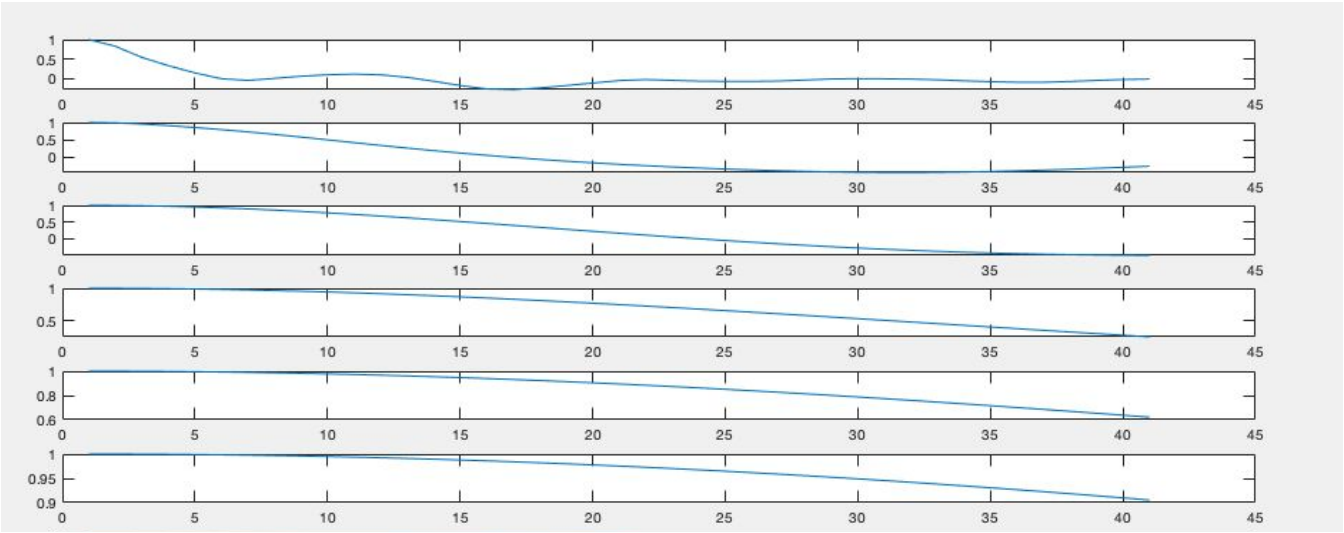
Result:





Separation of Unvoiced and voiced signal using a threshold





Autocorrelation functions