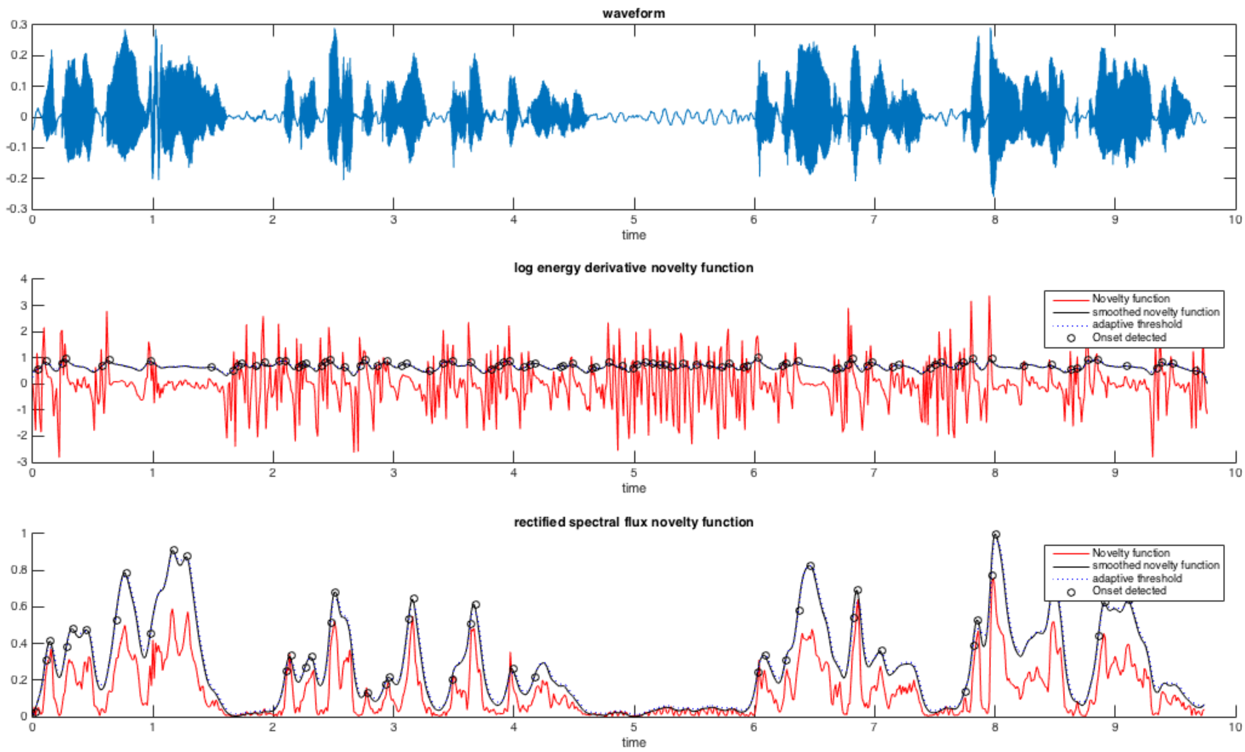
ASSIGNMENT 2

Anmol Monga
N18513543

ANALYSIS

Q1: female_short.wav

a)

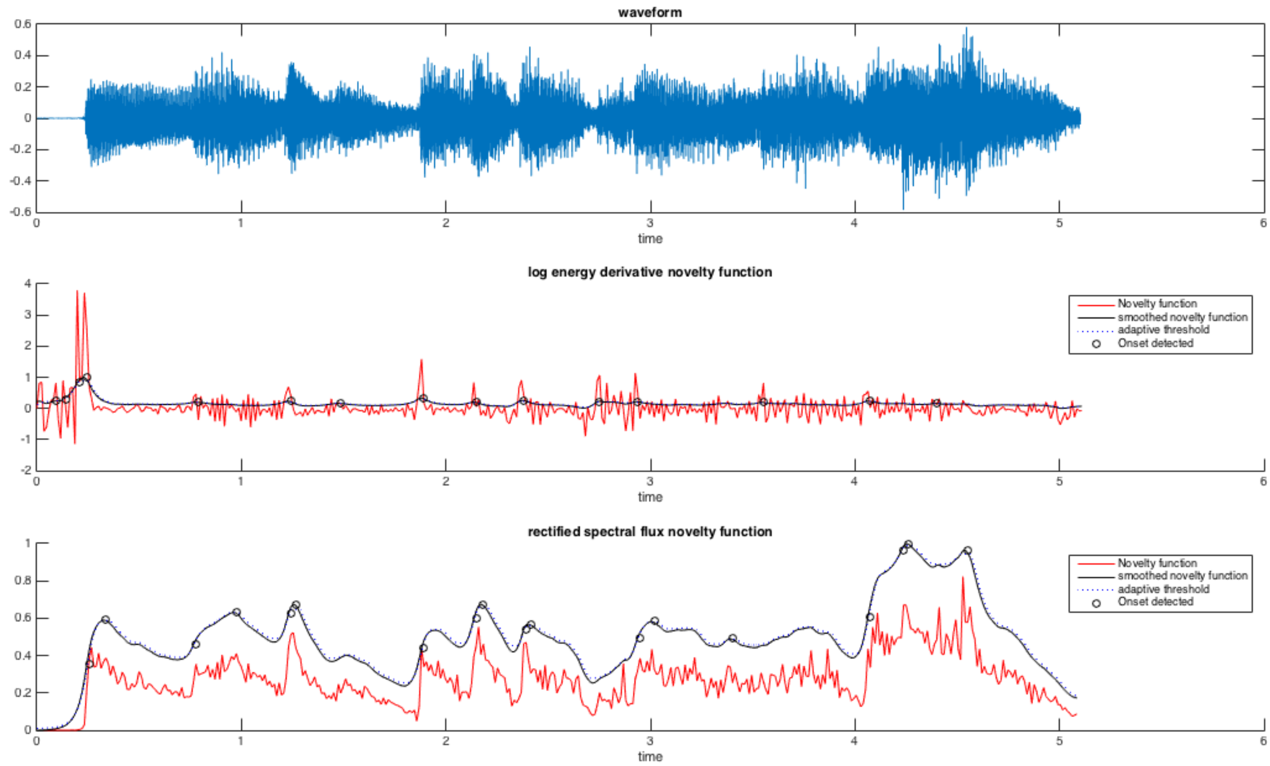


a) The Sound in this section is a female voice. Which has a pronounced onset, transient, steady state and decay. They gap between each attack makes the process of novelty detection effective.

b) The first plot Is of the sound wave. The second graph shows the log energy derivative novelty function. as we can see the novelty function obtained from log energy derivative novelty function is able to detect onset function which has low magnitude as easily as onset function of high magnitude. The log energy derivative novelty function is able to detect low noise level because it takes into account the noise level of the signal in comparison to the background noise level and hence is able to detect instances of attach for even low noise level.
The second plot on the other hand is the rectified average spectral change in the spectrogram of the first plot. It can only indicate change in spectral energy. It cannot not indicate any change in notes or onset of any string instrument as it involves a lot of change in harmonically related notes. As we can see from the plot the spectral flux indicates a lot of the offsets but if the offset is low in amplitude or is close to large energy sound burst the novelty function is unable to indicate the onset. In case of Log energy derivative function there is a lot of over-estimation, In this example rectified spectral flux estimation has done a better job.

Pianno.wav

a)

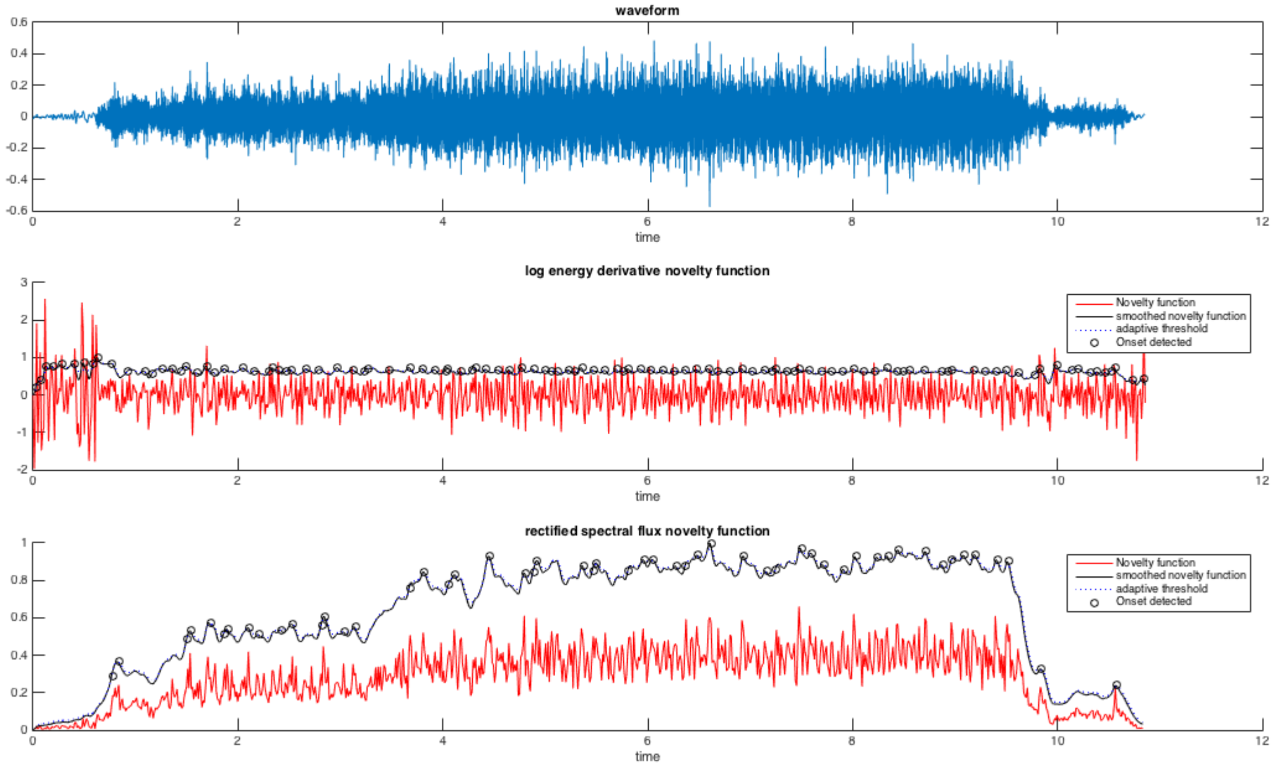


b) The above sound wave is a percussive sound . In Piano the noise is generated by a needle hitting a string and thus each onset in indicated by burst of energy in the time domain. Hence temporal detectors do a great job in finding the burst in energy.

c) The log energy derivative function is a temporal onset detector. Since Pianno generates a percussive sound. We can see log energy derivative does a good job detecting the onset when compared to rectified spectral flux novelty function. The log energy spectral flux can indicate even low energy burst when compared to the rectified spectral flux. Further we can see that rectified spectral flux novelty function has done a lot of over estimation of the onsets.

Sewing-machine_short.wav

a)



b) This a sound of a sewing machine, which is contains uniform burs of noise which are fused together with each other and appear as a single burst when we look at the plot of the waveform of the sound of sewing machine.

c) Both the log energy derivative novelty function and rectified spectral flux novelty function do a poor job of detecting the onset. The log energy derivative is able to detect offset in the low energy burst too. But overestimates all the onsets of the bursts by a large margin.
For the rectified spectral flux function, it is able to estimate the onsets but misses a lot of the onset. It is not able to detect onsets in low energy burst.

Q2:

a) Win_size: This is the window size . for the log energy derivative the large hamming window function means a smaller lowpass cutoff hence a more smoother signal. If the novelty function is smooth the median is pretty close to the the smoothed novelty function and hence lower number of onset detected and vice versa is true.
   In case of spectral flux novelty function, lower window size with proportional decrease in hop size means that the resolution is high compared to initial state and hence a smoother transition and thus closer to median and hence lower onset detected which will increase the false negative but decrease false positive.

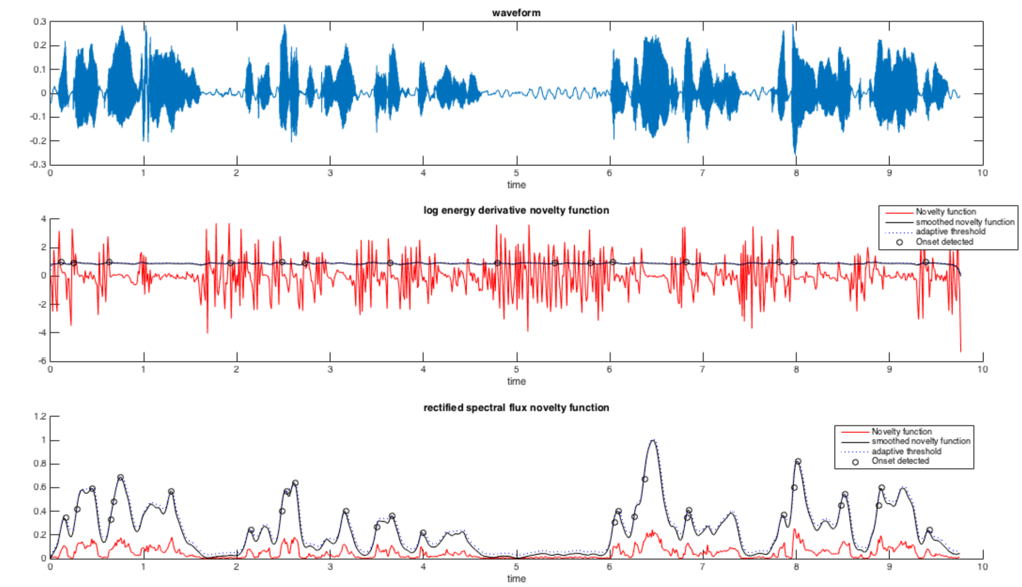
   Hop_size: the resolution of the novelty function for both spectral flux and log energy derivative decreases hence we can miss a lot of onsets which are pretty close to each other.

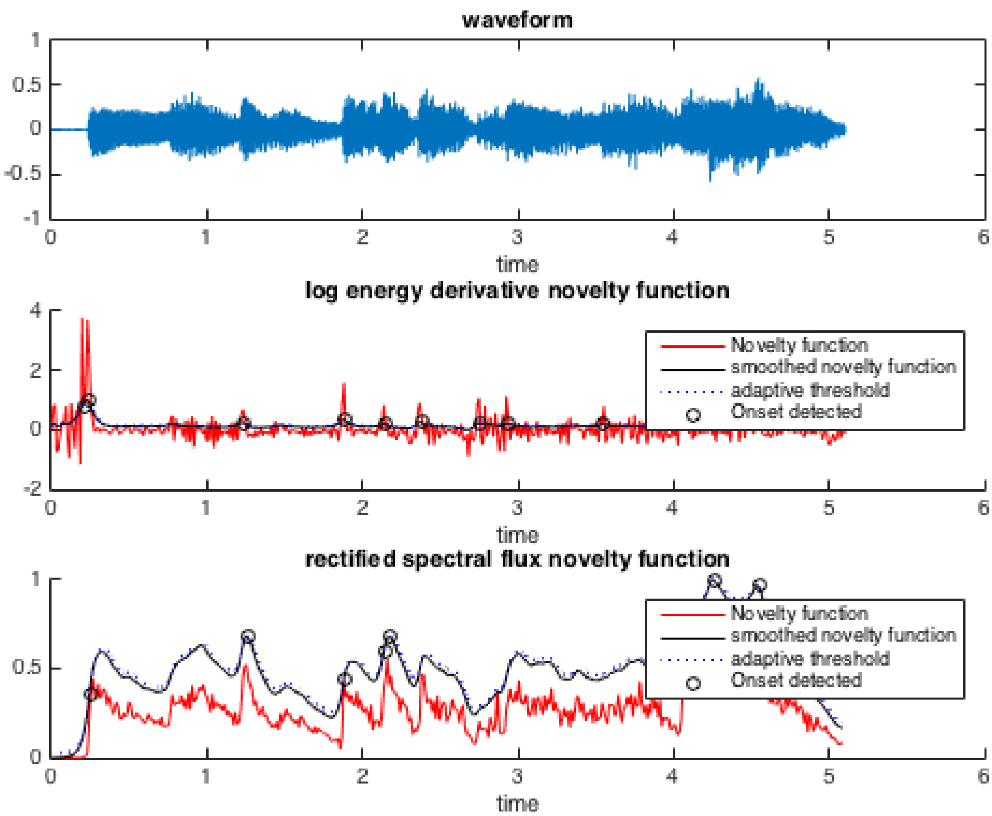
   Win_size: This specifies the cutoff frequency for the filter to smooth the novelty signals. Due to low frequency cutoff we can loose a lot of onset in the novelty function hence a lot of false negatives. We can also loose a lot of false positives due to this too. Hence we need to reach a perfect w_c for a very good onset.

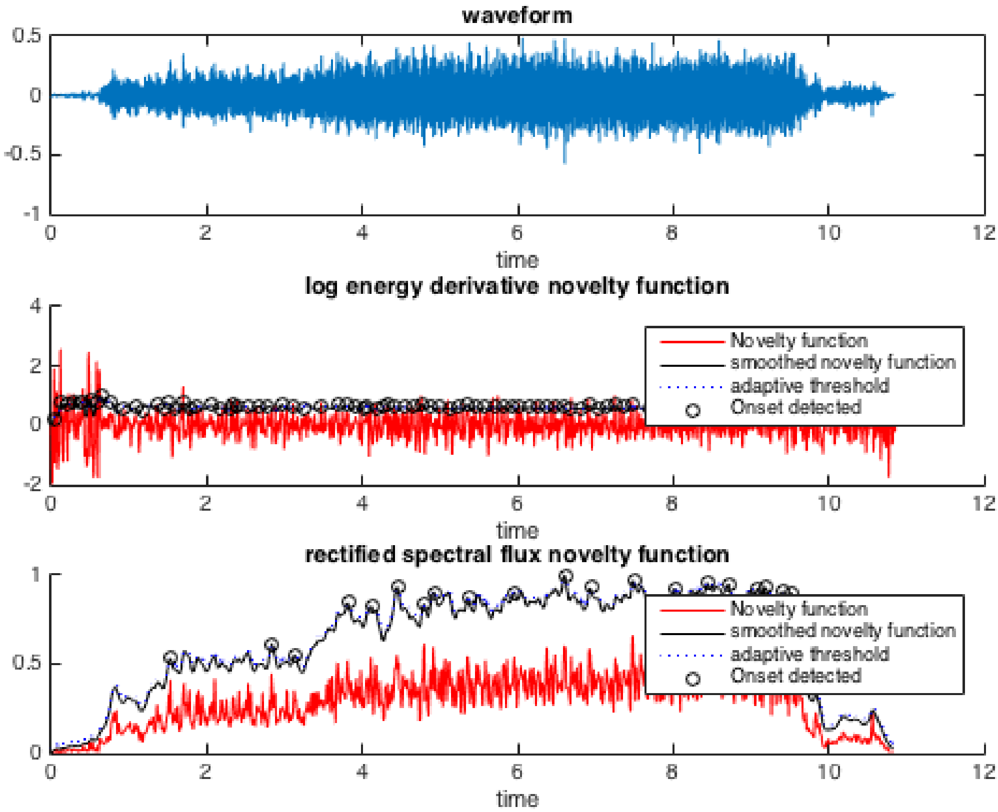
   Medfilt_len: This is the size of median filter used for adaptive thresholding. As med filter length becomes small it approaches closer in value to novelty function. But if length is increased it becomes less correlated to novelty function. For higher value of length we can detect most of the onsets on the higher value of sound but we miss a lot of onsets in the tuffs of the signals.So depending on how the novelty function looks like we need to decide median filter length is.
   Offset: offset specifies the offset added to the median filter of the adaptive thresholder. Higher level of offset, higher the threshold and hence very low false positive but high false negative. For lower value of offset high false positive but low false negative.

b)

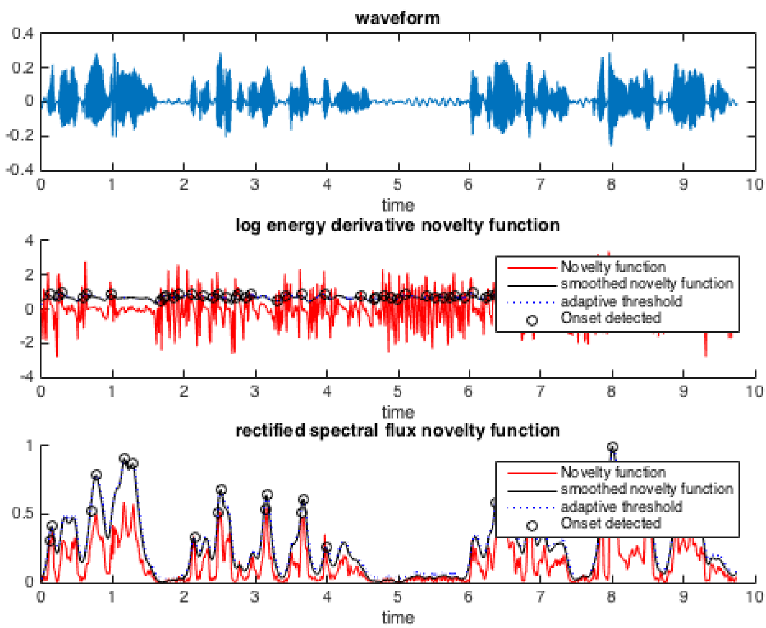
w_c = 3hz, medfilt_len = 10, offset = 0.016, win_size = 1024, hop_size = 512

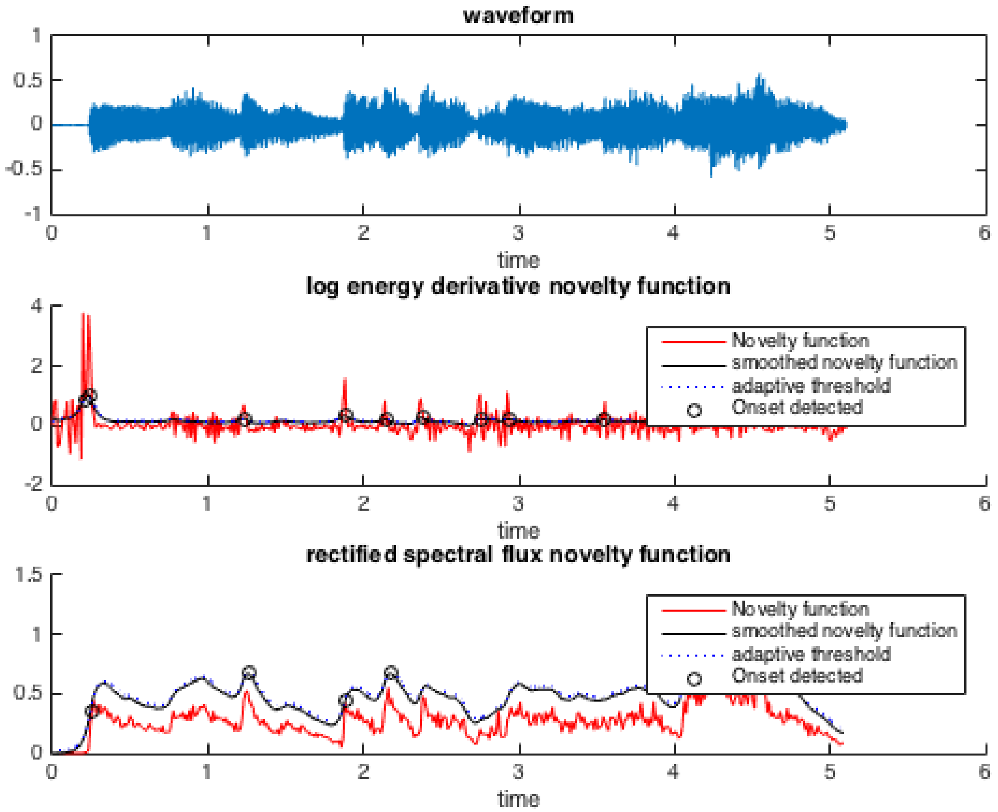




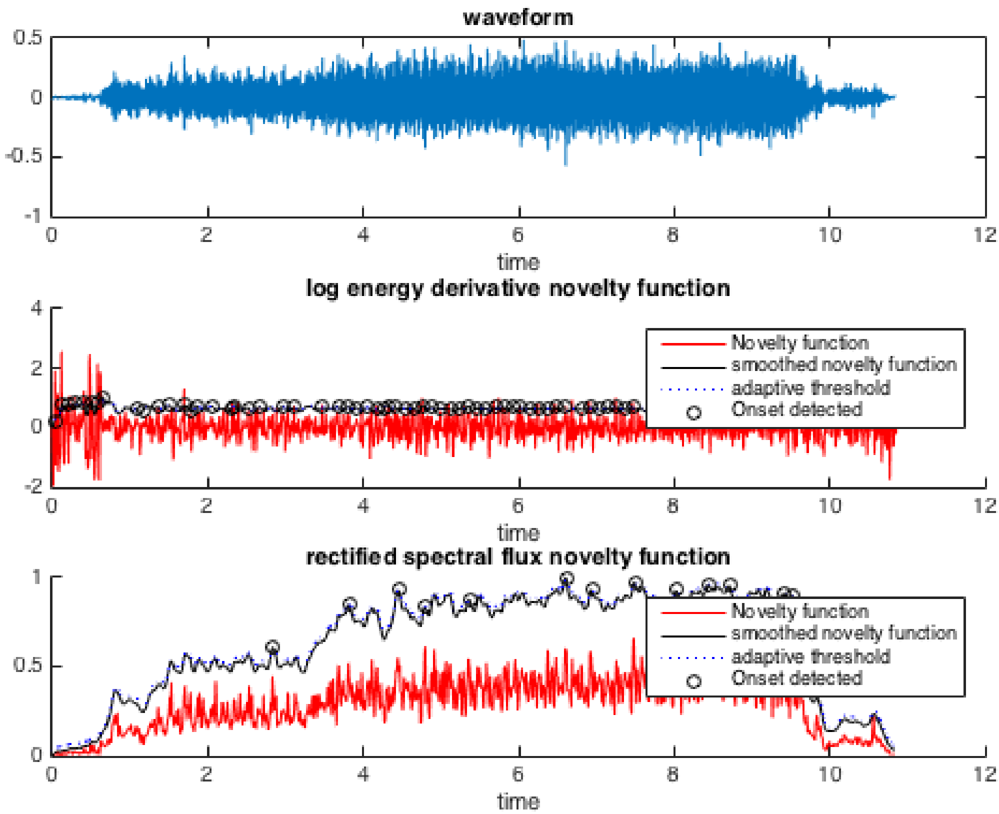


w_c = 4hz, medfilt_len = 8, offset = 0.026, win_size = 1024, hop_size = 512



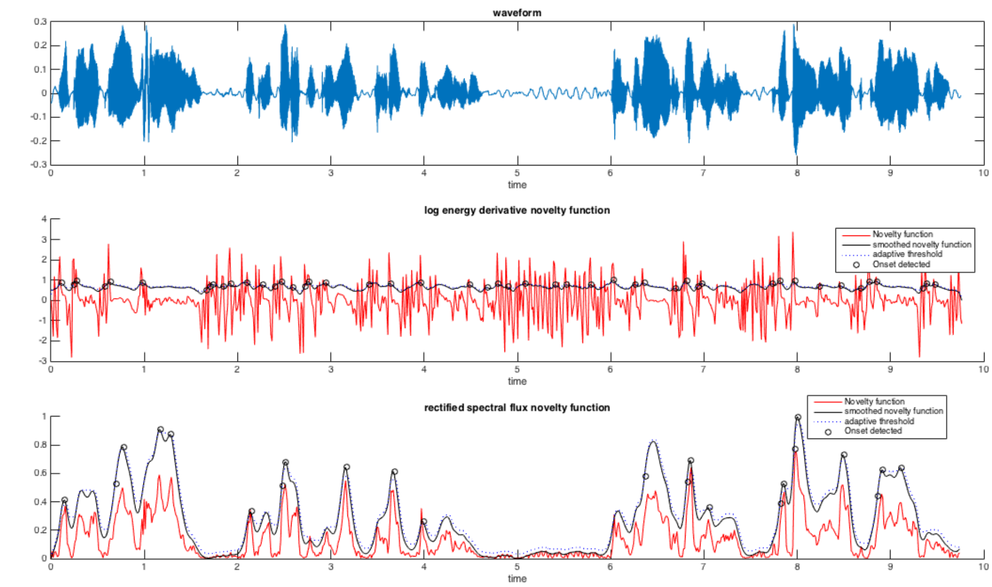


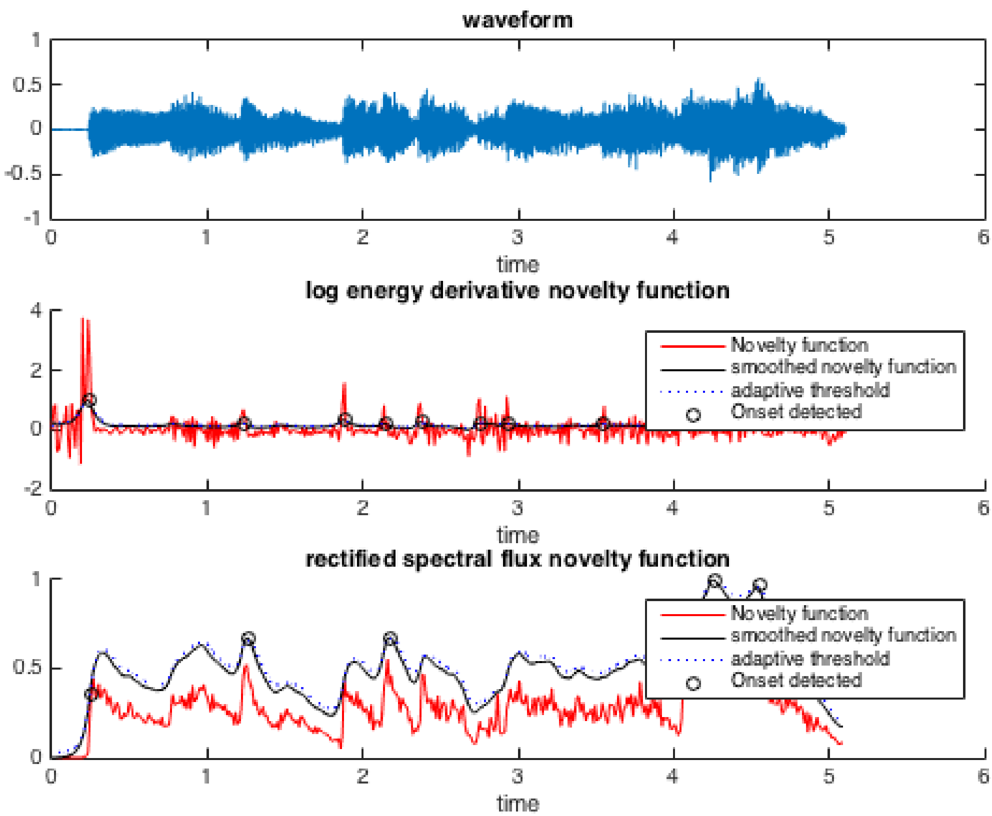
waveform
time
log energy derivative novelty function
Novelty function
smoothed novelty function
adaptive threshold
Onset detected
time
rectified spectral flux novelty function
Novelty function
smoothed novelty function
adaptive threshold
Onset detected
time

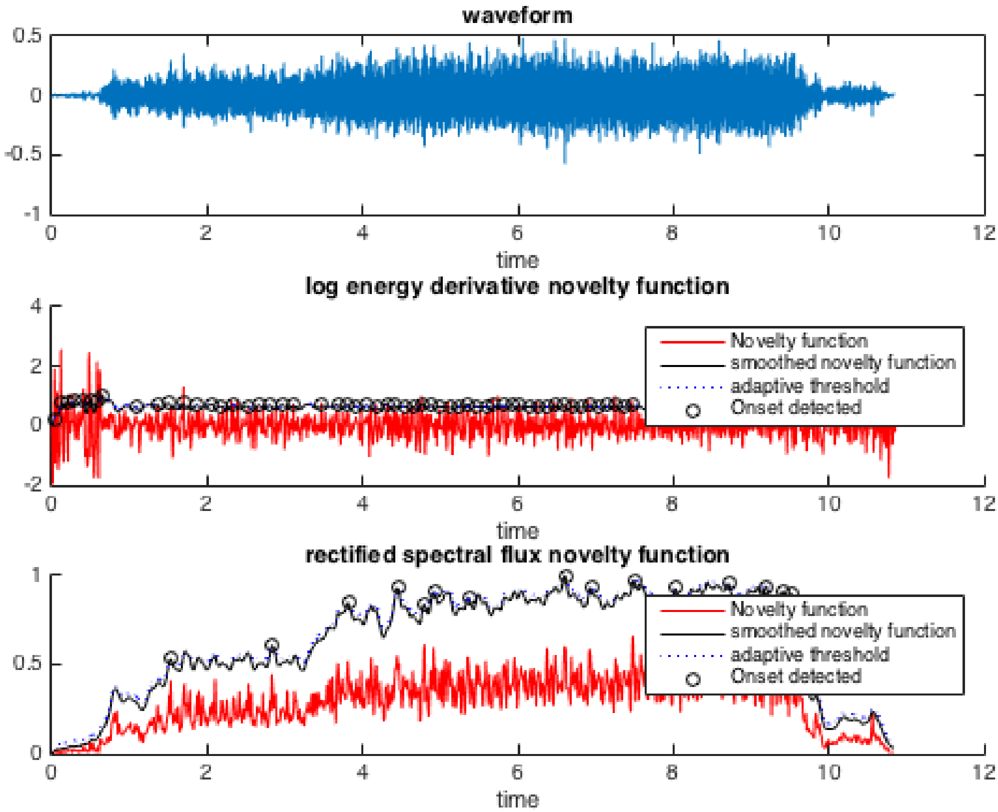


waveform
time
log energy derivative novelty function
Novelty function
smoothed novelty function
adaptive threshold
Onset detected
time
rectified spectral flux novelty function
Novelty function
smoothed novelty function
adaptive threshold
Onset detected
time

w_c = 3hz, medfilt_len = 10, offset = 0.03, win_size = 1024, hop_size = 512






waveform
0.5
0
-0.5
-1
0
2
4
6
8
10
12
time
log energy derivative novelty function
4
2
0
-2
Novelty function
smoothed novelty function
adaptive threshold
Onset detected
time
rectified spectral flux novelty function
1
0.5
0
Novelty function
smoothed novelty function
adaptive threshold
Onset detected
time