# Assignment 4

Amruta and Rithesh

1. Self-Distance Matrix:
   - MFCC SDM has a prominent feature visualization: the different segments are much more easily visible.
   - The chroma is very similar to MFCC SDM, but it is much less pronounced.
   - Chroma aggregates everything, so it is difficult to see the segments than it is in the MFCCs. This is because the song has a relatively flat/uniform spectral distribution.
   - The horizontal and vertical stripes tell us how much a given time segment is similar / dissimilar to other segments of the song. If it is yellow or brighter, then it is very dissimilar. Blue, or a darker shade indicates it being similar. For example, in the MFCC sdm we see yellow vertical stripe at approximately 120-150sec which is dark blue only at the diagonal but dissimilar at other locations.
   - The stripes parallel to the diagonal indicate sections being similar to one another. These could become a segment of the song. The 3 parallel lines in the chroma corresponding to 82 to 100 s repeat again in the segment 177 to 197 s and again from 309 to 333 s. This is the chorus of the song!
   - Blocks along the diagonal:
     - 1 → 10-30 s
     - 2 → 50 - 100 s
     - 3 → 105-130 s
     - 4 → 150 - 200 s
   - Blocks 1 and 3 are similar, as apparent from the Chroma SDM. They correspond to the Main Riff and Verse.
   - Blocks 1 and 4 are similar (infact the whole time segment can be seen to repeat). These sections are the bridge+chorus of the song.
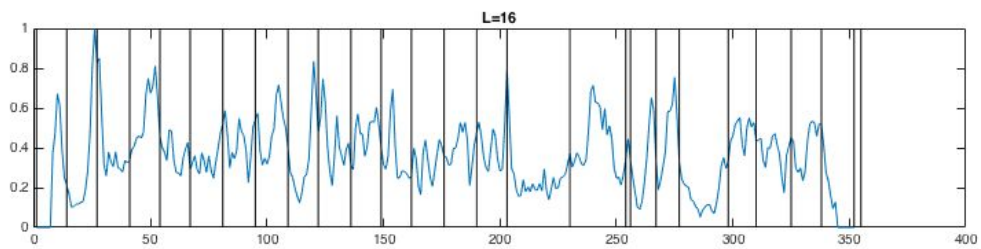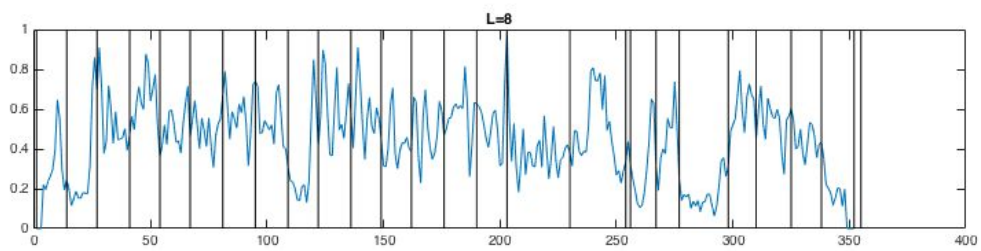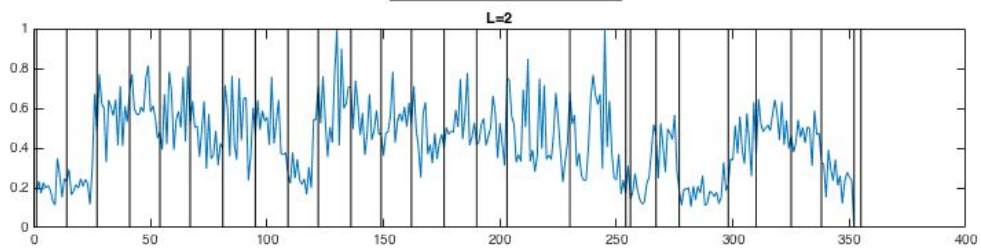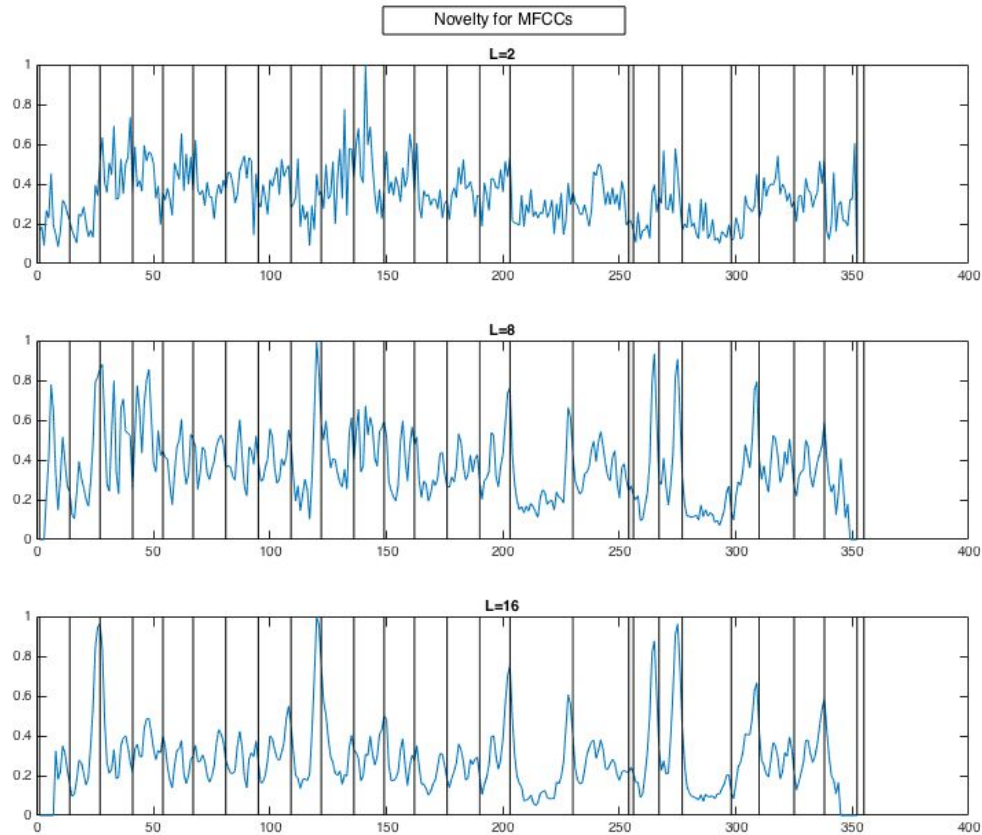
Chroma

MFCCs

MFCC

Chroma

MFCC
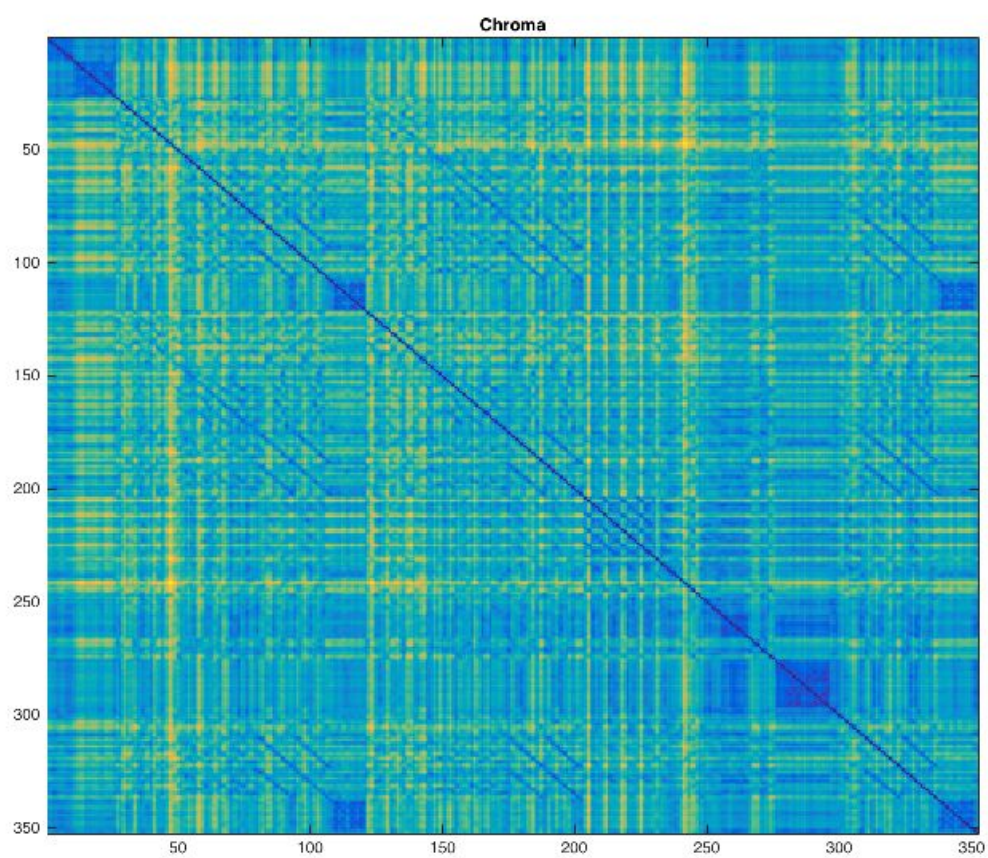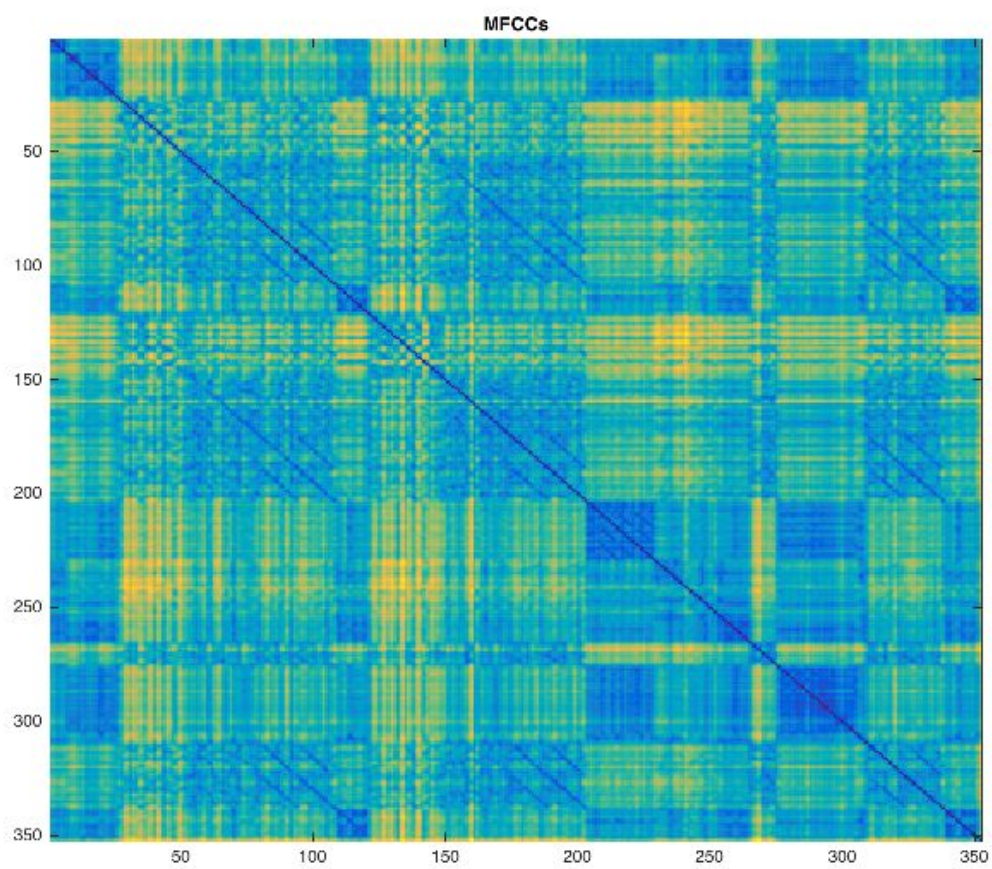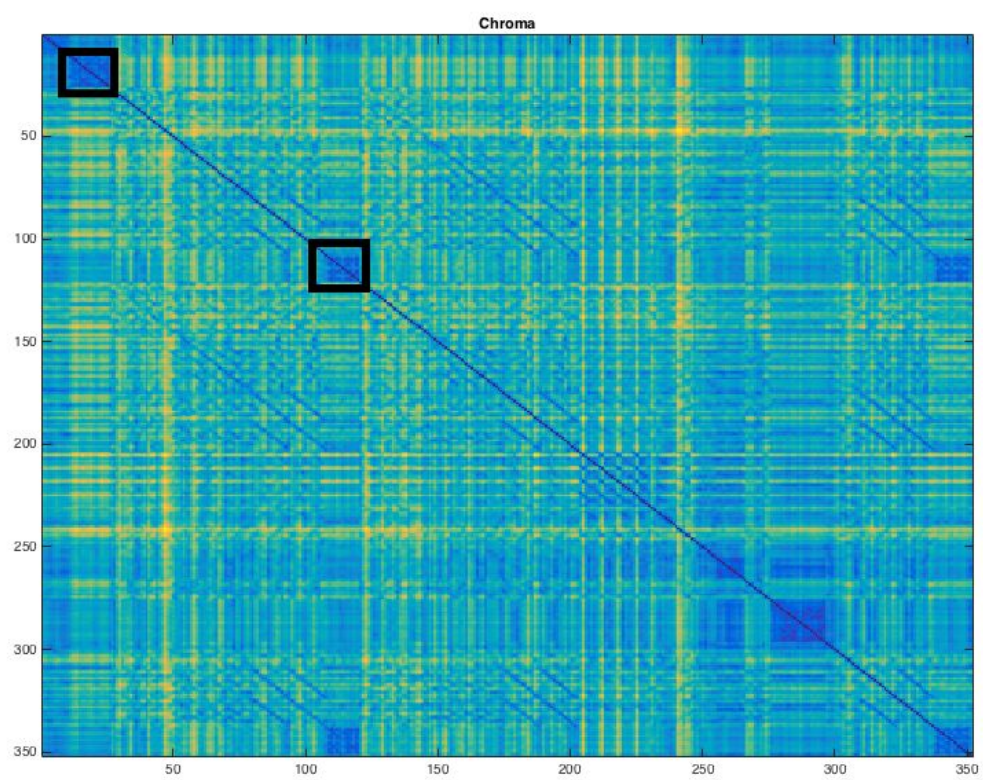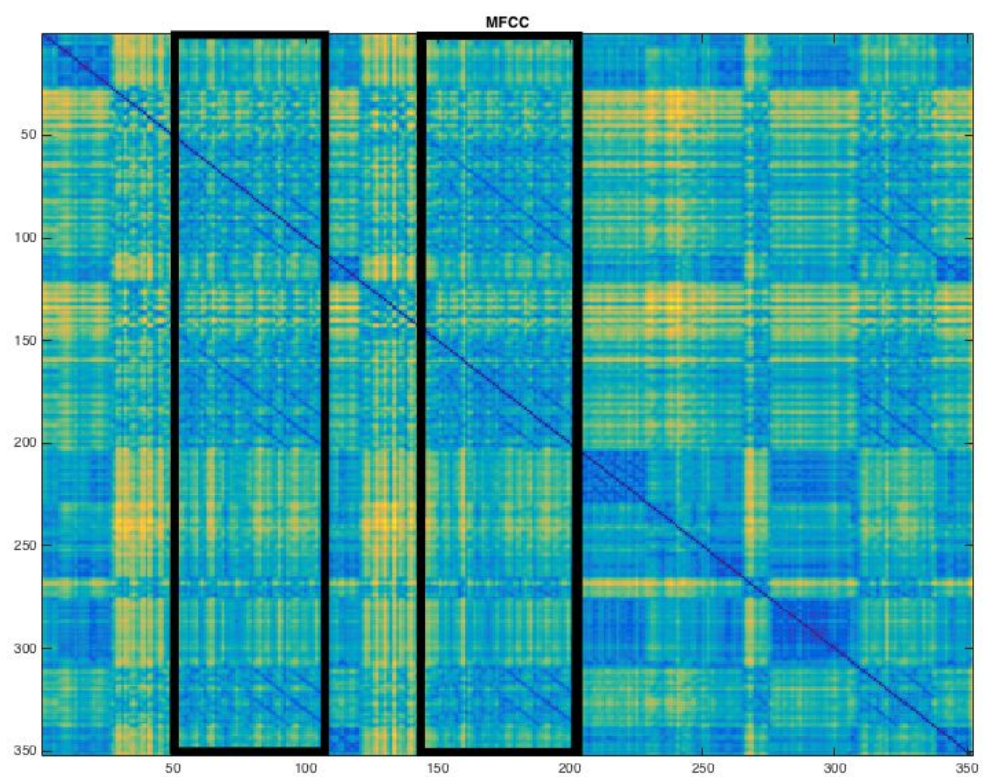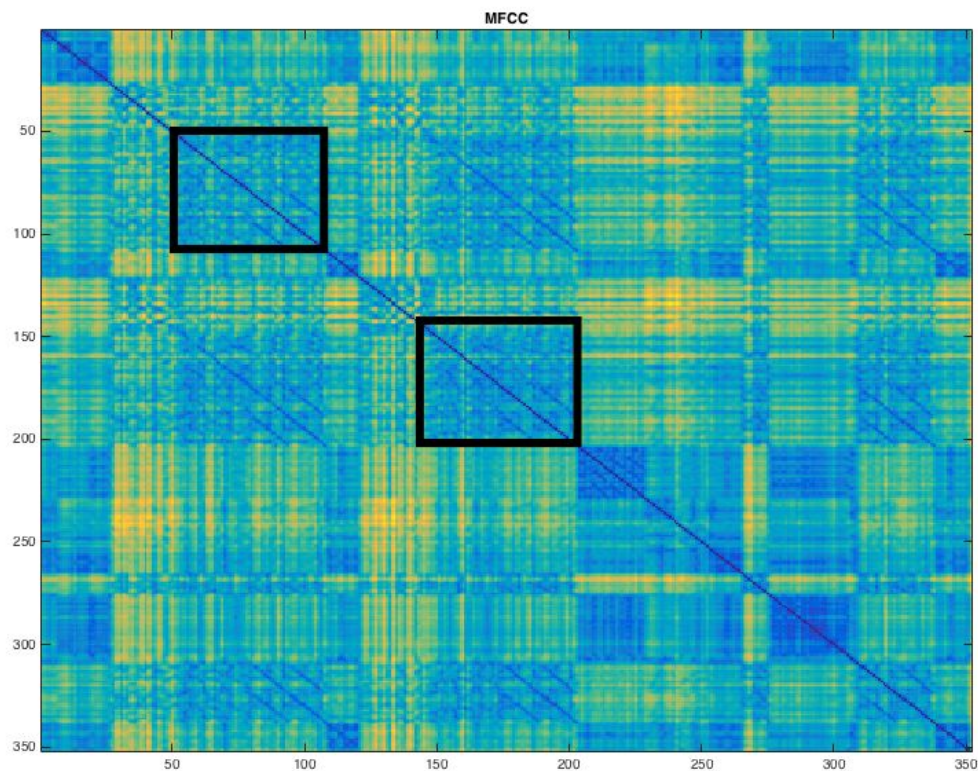
2. CheckerBoard Kernel Novelty Function:
   ● L = 16 gives a better result. This is because the novelty plot looks smooth
     with clear high values in location of ground truth and low values in the location
     where ground truth is not present as seen in figure of novelty for Chroma or
     MFCCs in the region 200s-225s there are low values for L=16 as opposed to
     L=8. The relationship between the ground truth and the novelty is that the
     novelty represents section change.
   ● It computes the difference between the values along the diagonal and its
     surroundings. So, if there is a difference ('onset') in value along the diagonal,
     it outputs a peaky value. This is is essentially a HPF.
   ● Observing the SDM and the novelty shows that MFCCs are performing better
   than Chroma.

Novelty for Chroma

Novelty for MFCCs

3. Lag-Distance Matrix & repetition detection
 - **computeBinSDM**: For MFCC, the best threshold is 0.35, for a normalized SDM. The mean is 0.4339 and was the best representing the song segments, along the diagonals. The representation for chroma was not very informative, however, the parallel diagonals, as apparent from question 1 were still notable, for a threshold of 0.4. The mean for the Chroma SDM values was 0.5266.
 - The enhanced lines give us the extracted regions. The start and end time of these repeating extracted regions highlight a segment of the song. From the annotation we infer that we have detected the chorus of the song.
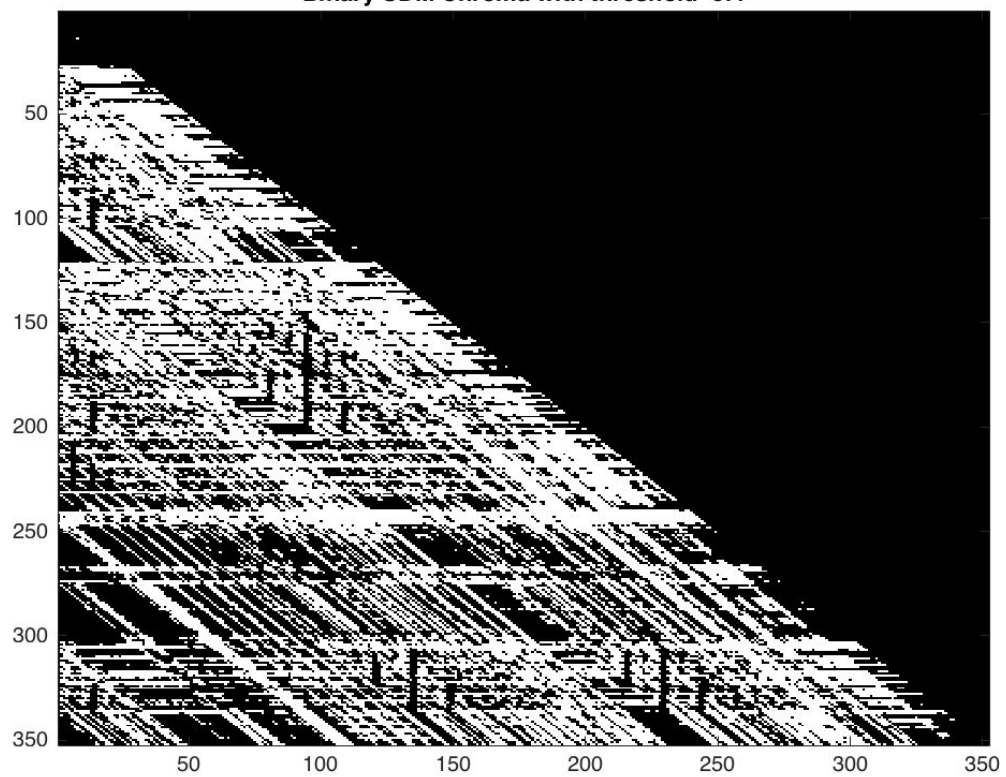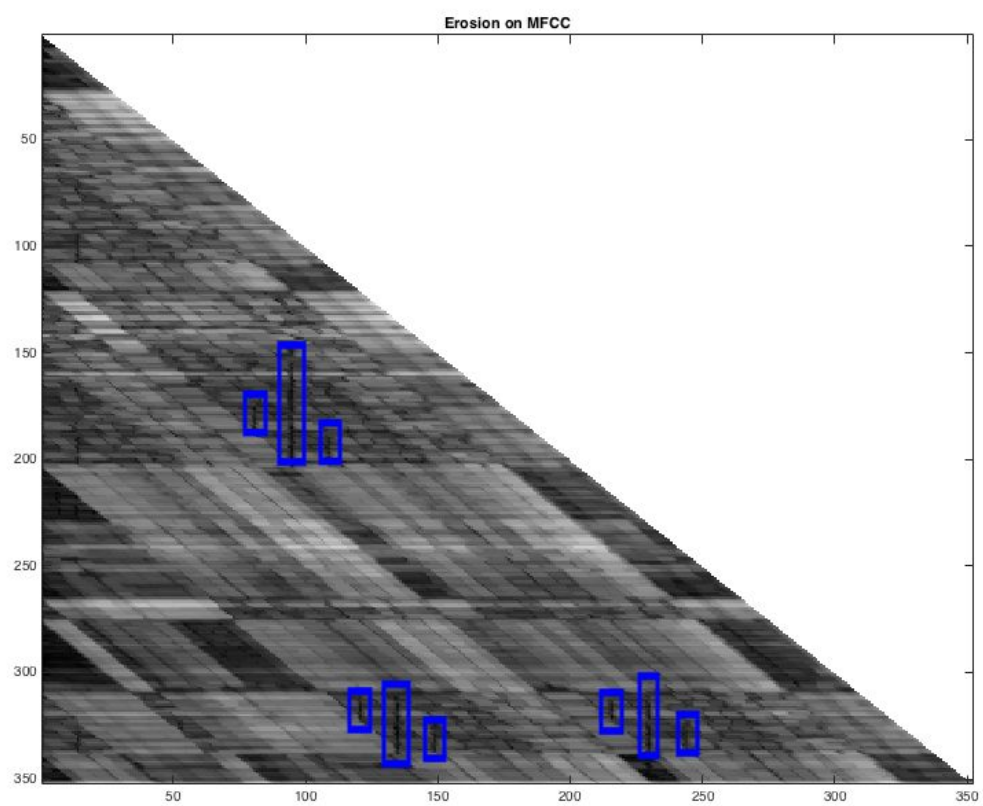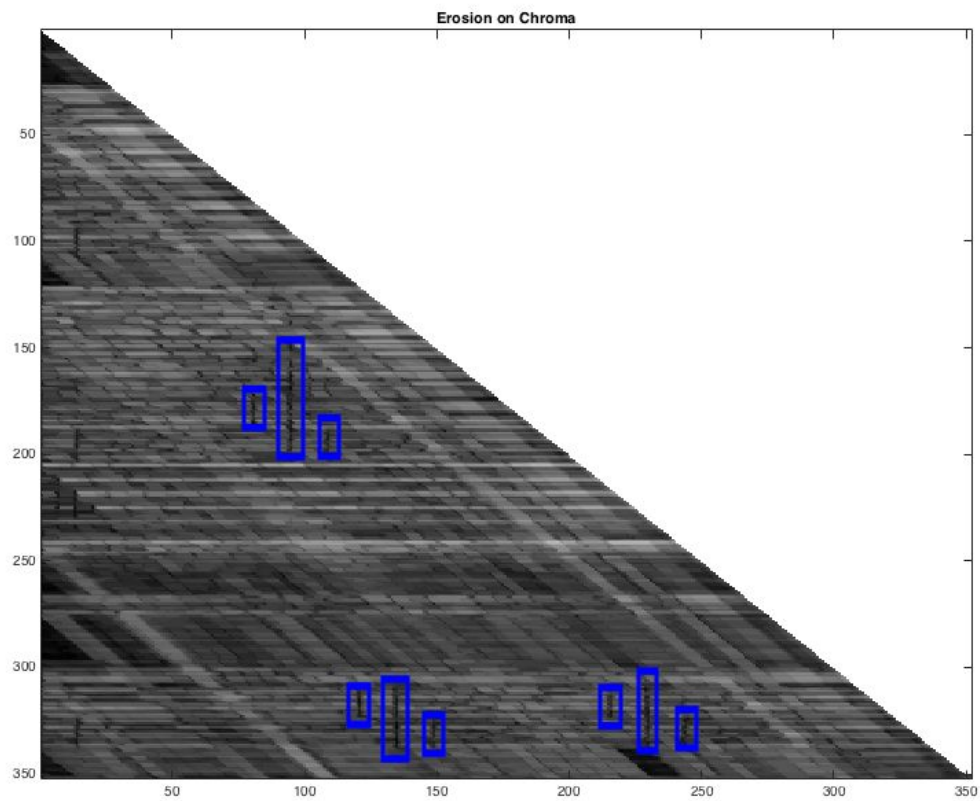
Lag Distance Matrix with Chroma

Lag Distance Matrix with MFCCs
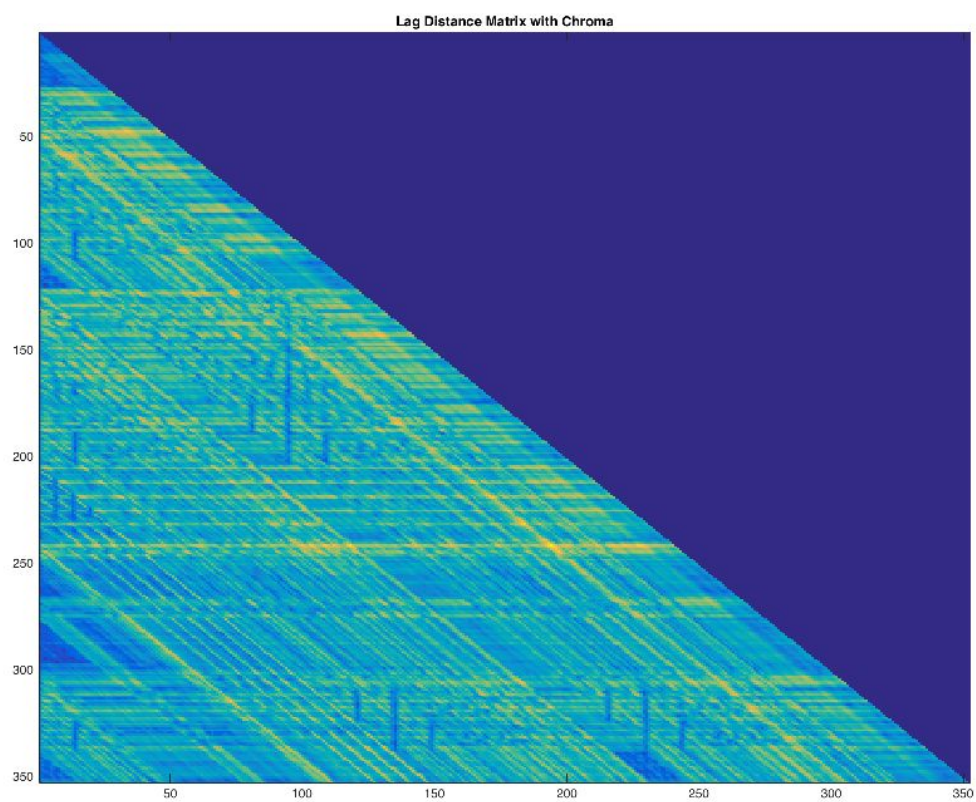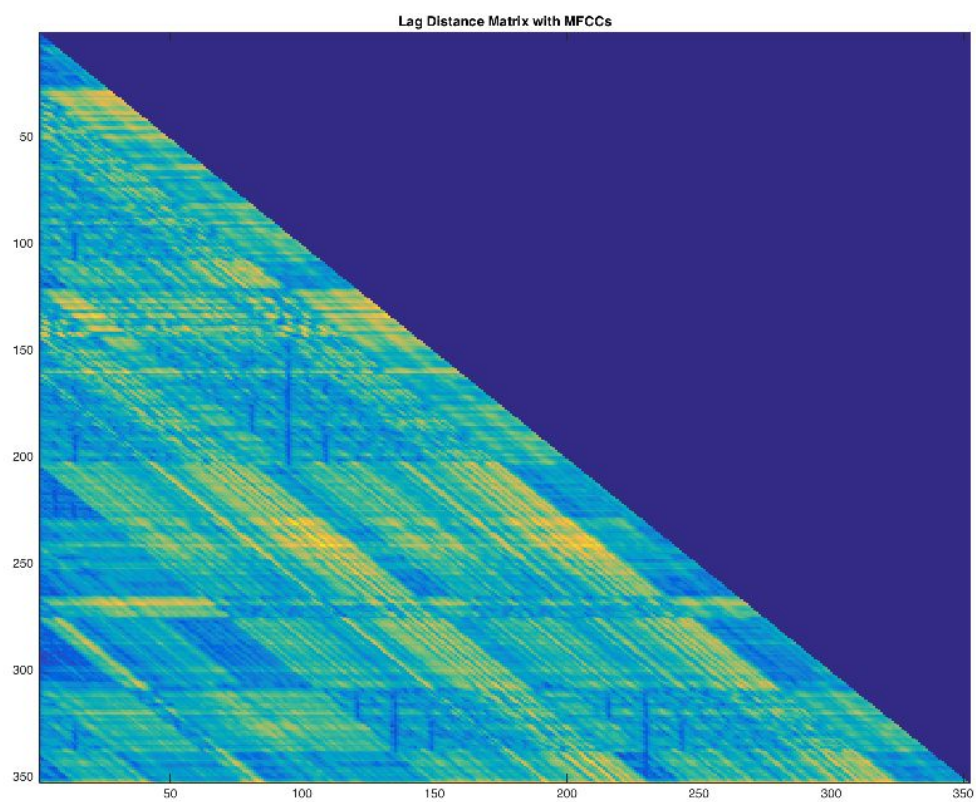
Binary SDM MFCC with threshold=0.35
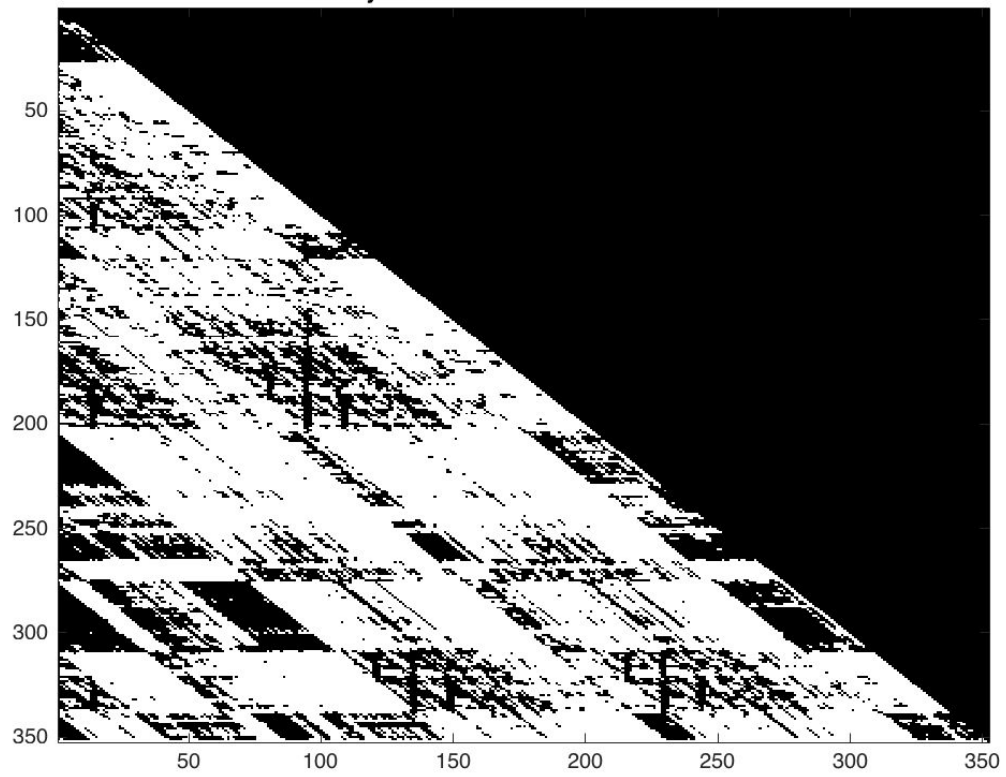
Binary SDM Chroma with threshold=0.4

Erosion on MFCC

Erosion on Chroma

4.Bonus:

Using the SDM technique, we can infer the segments of the song. Now, after we extract the segment, we may apply the same technique on the segment to get the sub-segments. Sometimes, like in our case, we see that the repeating segments consists of three parallel lines. These three parallel lines could be treated as markers for the sub-segments. The smaller length lines could also tell us about the sub-segment structure.