MMT - Assignment 3

Sreeja Guduri - 2021102007

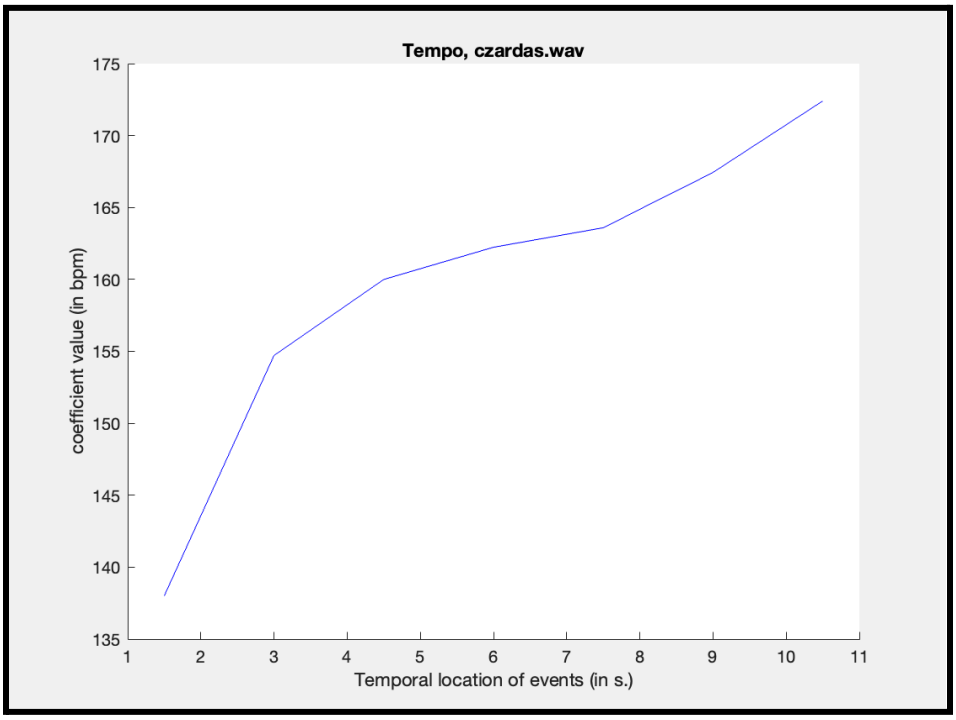
Rhythm and meter

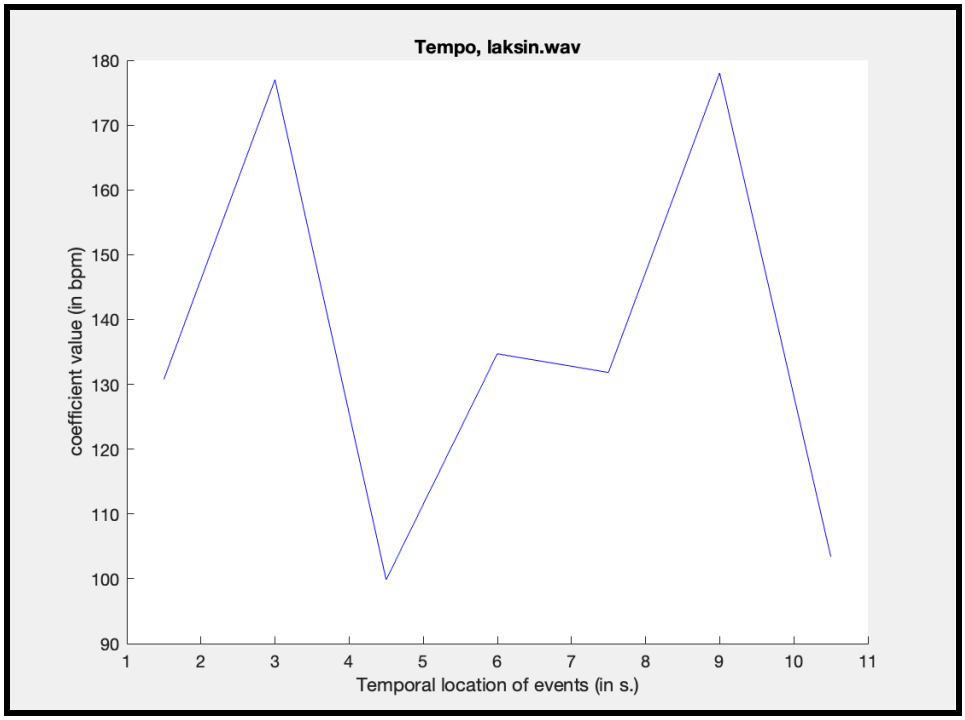
1.

Songs	Perceptual tempo (bpm)	Calculated tempo (bpm)
ragtime.wav	140	129.471
vivaldi.wav	180	182.4698
valse_triste_happy.wav	110	121.5723
laksin.wav	92	95.32
czardas.wa	145-185	162.8764

- The perceptual and calculated tempo estimates, while mostly similar, did show some discrepancies.
- One reason for the difference could be that we tend to calculate tempo using the most emphasized beat (most perceptually distinct). However, the function internally calculates tempo using an autocorrelation function for different tempo ranges - and choosing the one that fits the best. Thus, this could lead to variations because the most distinct beat may not represent the song very well.
- Also, as noted - the 'laksin.wav' and 'czards.wav' files have different tempos in different parts of the song and thus, the calculated tempo could be inaccurate because it averages out all the estimated tempos.

2. Frame-based tempo analysis is used for the audios with variable tempo and the plots obtained are:





For ‘czardas.wav’, the tempo ranges from about 137 to 173bpm. This is roughly in agreement with what I estimated using the metronome, and the time-varying pitch can be analysed using the graph.

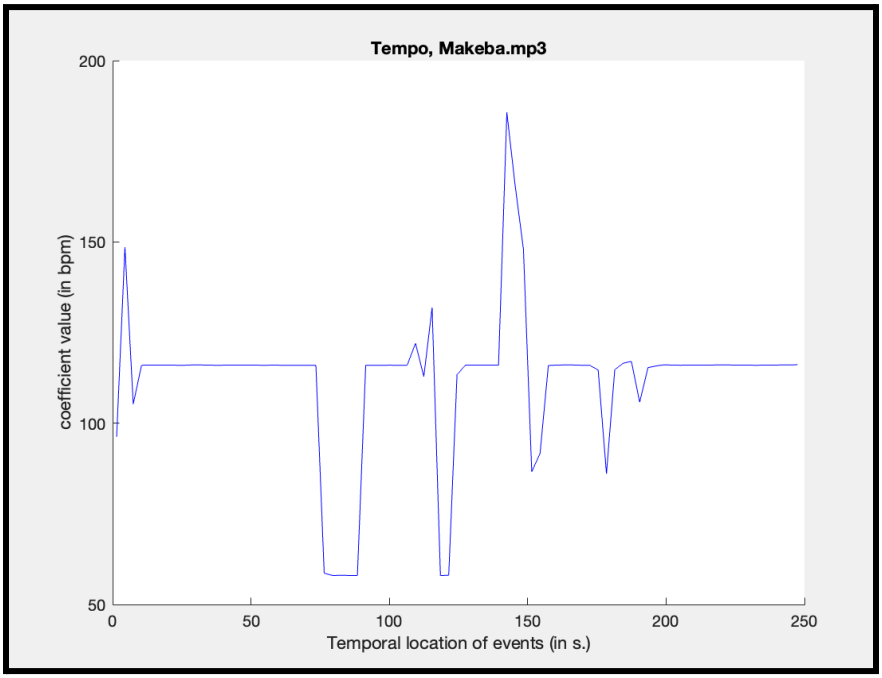
Similarly, for ‘laksin.wav’, the pitch ranges between 100 and 178bpm though I could not perceptually find this range of tempo difference using the metronome.

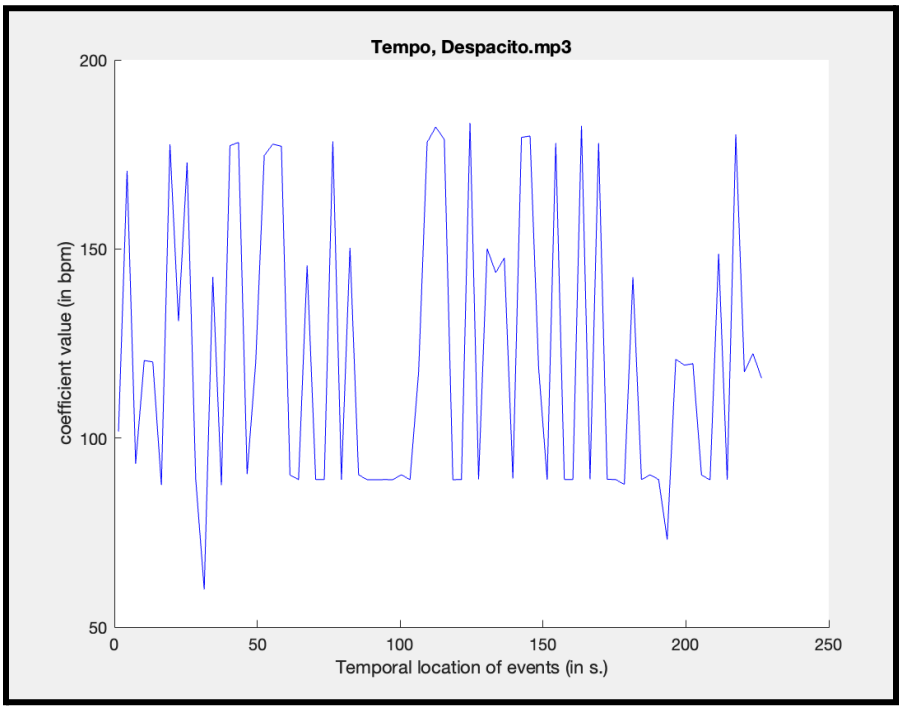
2.
Perceptual Tempo rating:
Rite_of_spring < Despacito < Makeba < Derezzed < Stream_of_Consciousness

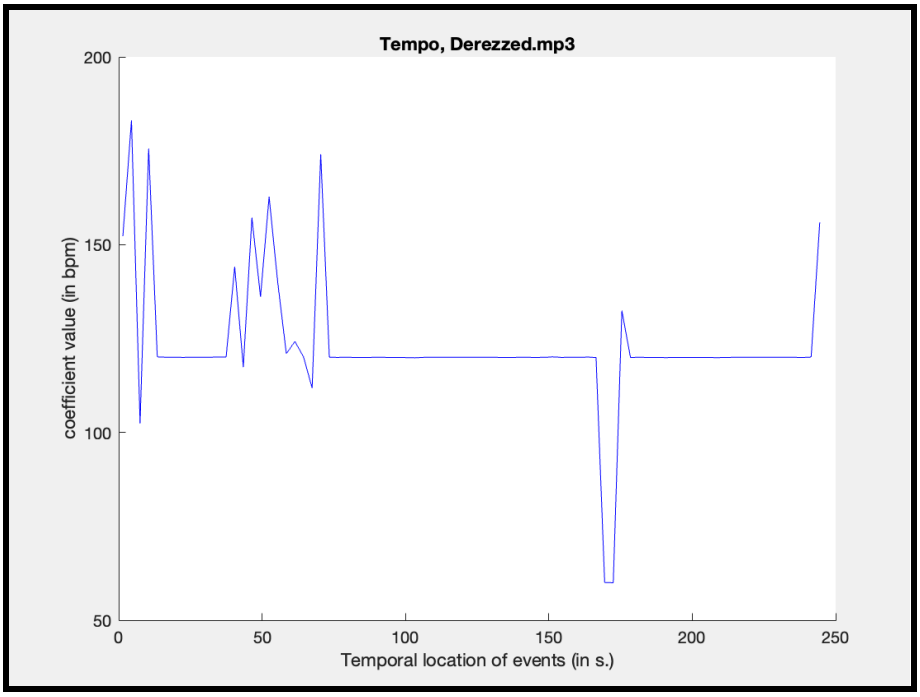
Computational Tempo rating:
Despacito < Rite_of_spring < Makeba < Derezzed < Stream_of_Consciousness

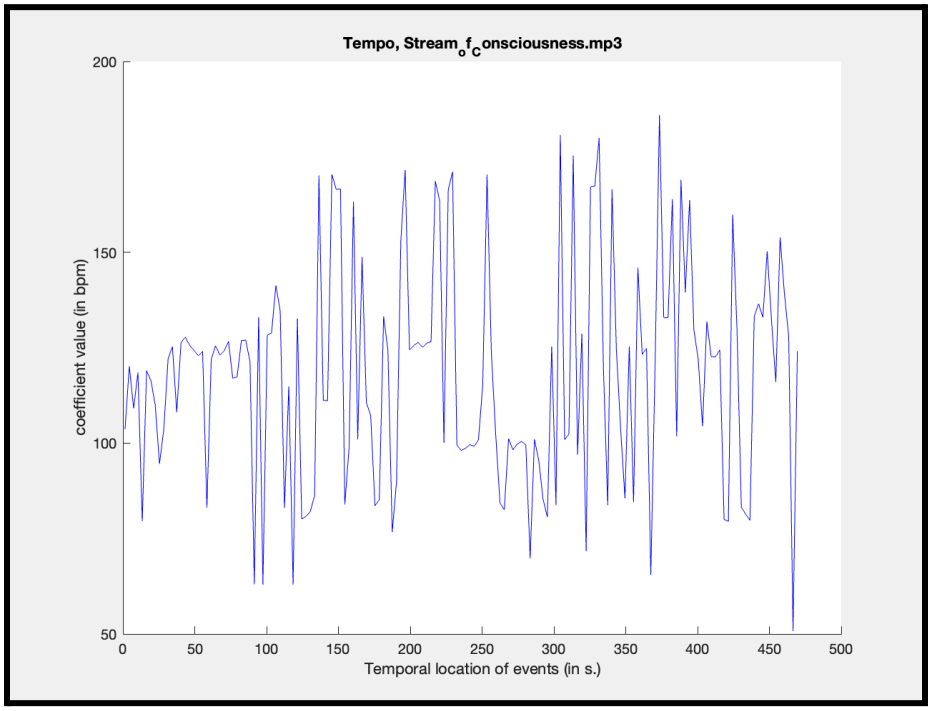
Thus, the perceptual ordering mostly agrees with the computational ones. Discrepancies could arise if the songs have variable tempos throughout (which is the case for most of these songs) - so the function will only return the average tempo.

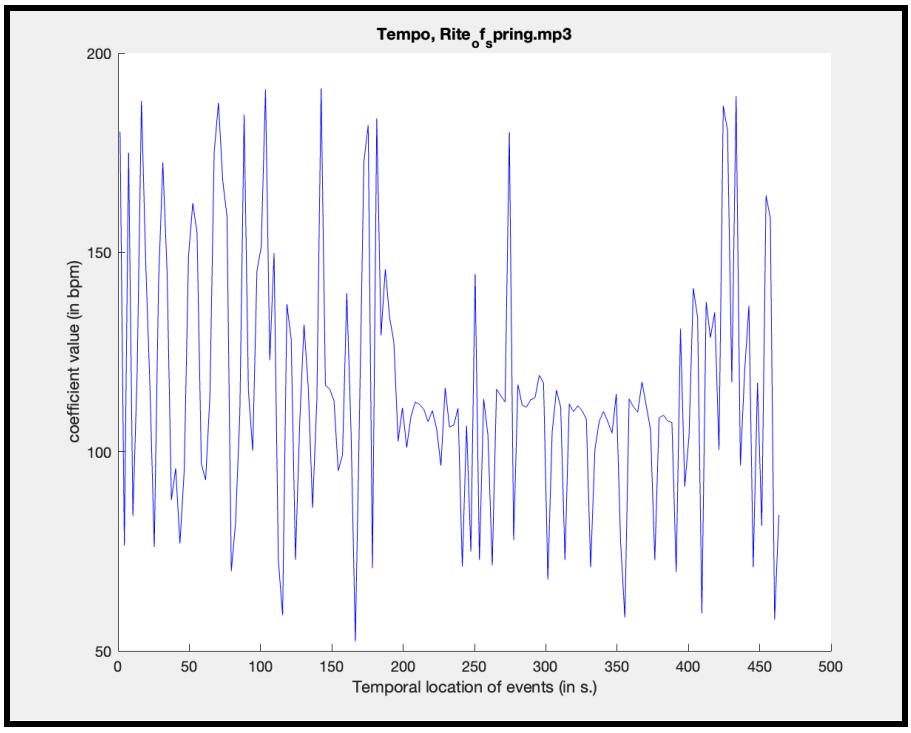
The evolution of tempo in the songs when plotted is:






Tempo, Derezzed.mp3
coefficient value (in bpm)
Temporal location of events (in s.)
200
150
100
50
0
50
100
150
200
250


Tempo, Stream of Consciousness.mp3
coefficient value (in bpm)
Temporal location of events (in s.)
200
150
100
50
0
50
100
150
200
250
300
350
400
450
500



From the plots we can tell that ‘Makeba.mp3’ and ‘Derezzed.mp3’ have the most stable plots, and thus the most steady tempos. A constant beat and tempo would generally indicate higher pulse clarity (easier to follow the beat).

3.

Perceptual Pulse clarity rating:

Rite_of_spring < Stream_of_Consciousness < Despacito < Makeba < Derezzed

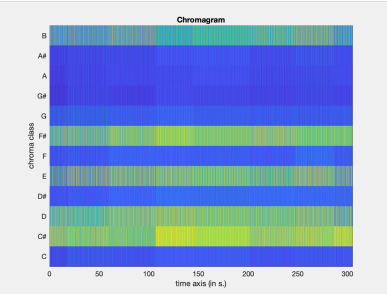
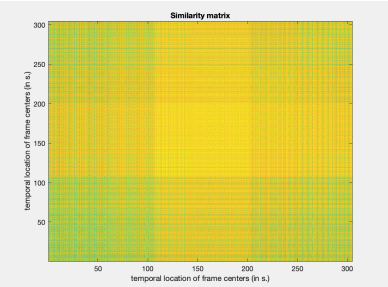
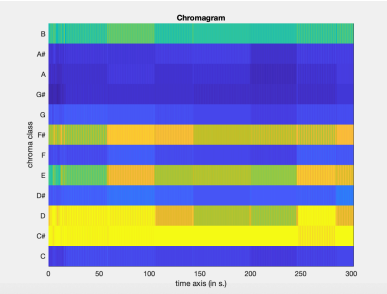
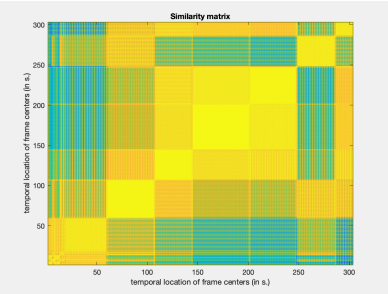
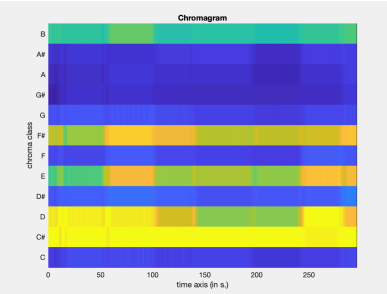
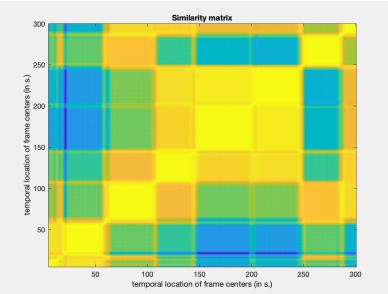
Computational Pulse clarity rating:

Rite_of_spring < Stream_of_Consciousness < Despacito < Makeba < Derezzed

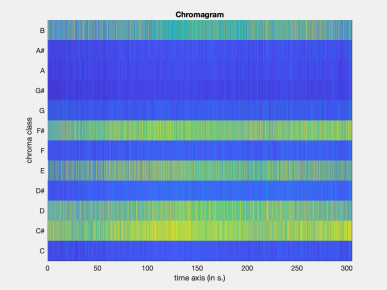
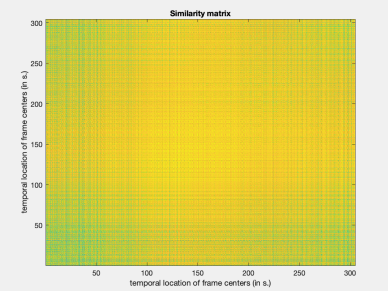
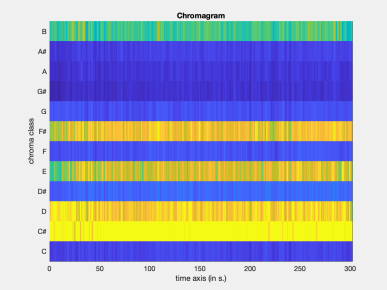
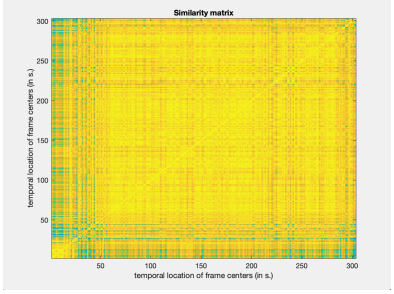
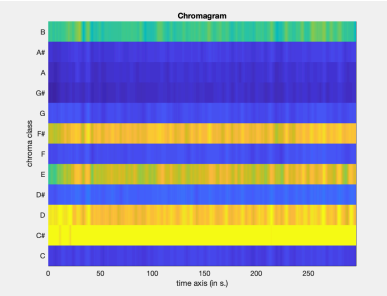
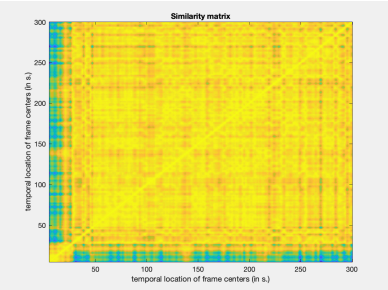
We notice that the pulse clarity ratings are similar, and we do notice the correlation between steady tempo and higher pulse clarity. Any discrepancies could once again be because the function algorithm is thrown off by multiple instrument timbres and a dynamic tempo.

Repetition in music

1.
For ‘Abrupt-trim.wav’:

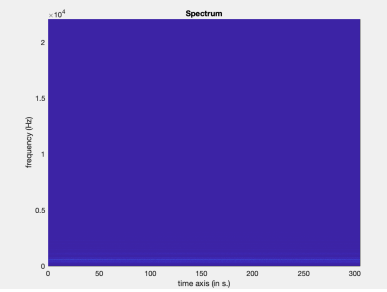
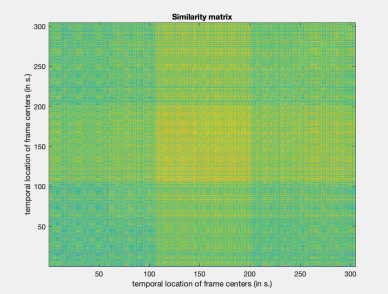
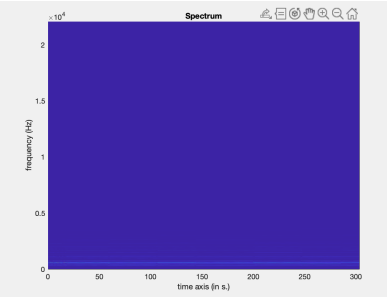
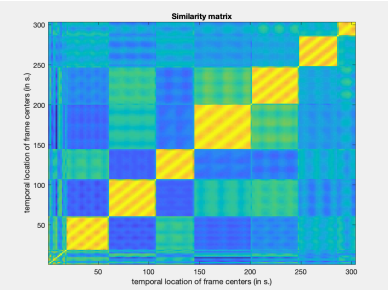
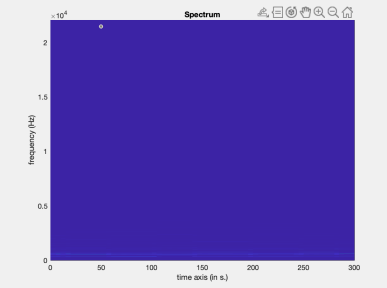
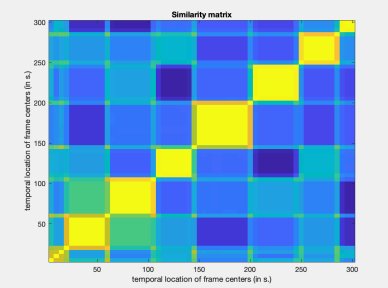
250ms		
3s		
10s		

For ‘Original-trim.wav’:

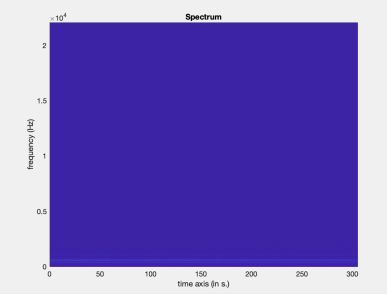
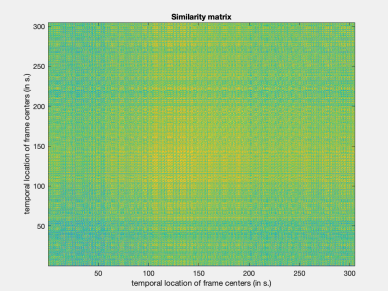
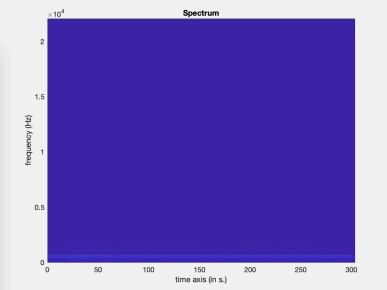
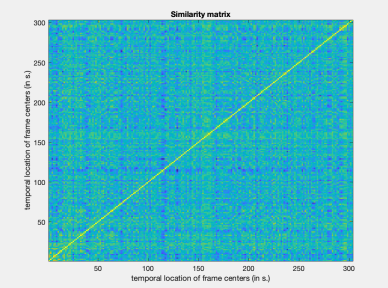
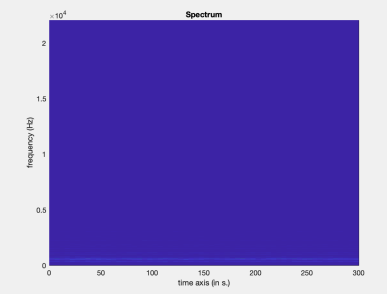
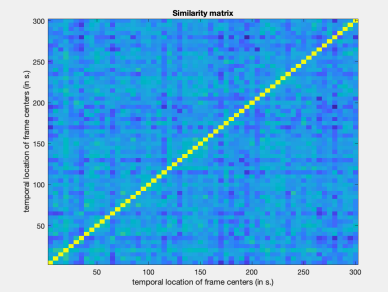
250ms		
3s		
10s		

2.

For ‘Abrupt-trim.wav’:

250ms		
3s		
10s		

For ‘Original-trim.wav’:

250ms		
3s		
10s		

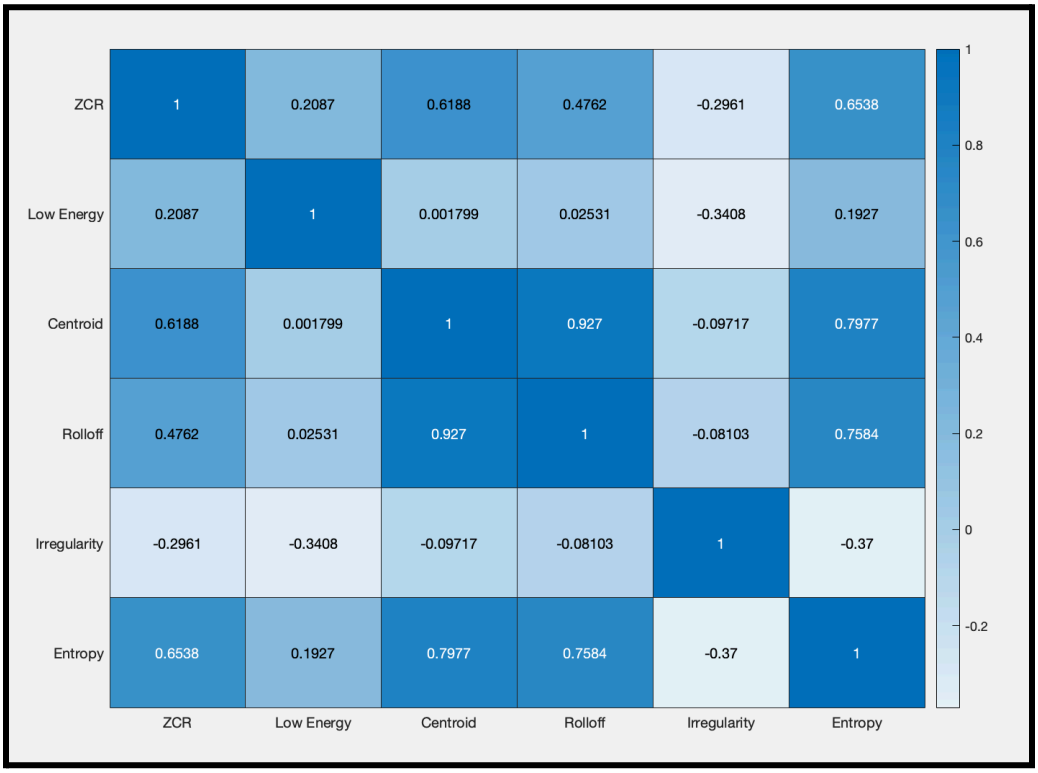
We notice that for both files, the granularity of the representations is the highest for a frame length of 250ms - owing to the high temporal resolution. A window length of 3s would best represent our perceptual understanding of repetition. The 10s frame length would be too long and would obscure out any of the finer musical repetitions, while a length of 250ms may capture too many fine variations. The 3s frame length provides a good balance between the two and is most similar to how we perceive music segmentation.

Timbre

1.

- Spectral centroid is related to how 'bright' the music sounds. Higher centroid indicates most of the energy is concentrated in the higher regions - contributing to the sound sounding more bright.
- Spectral rolloff is also a measure of where most of the frequency components of the sound reside. Higher rolloff indicates more high frequency content in the signal. This also correlates with how 'bright' a piece of audio sounds.
- Spectral entropy is a measure of how steady/fluctuating the frequency of the audio is. High spectral entropy indicates a constantly changing frequency. Thus, the perceptual correlate of this measure is how 'noisy' or 'tonal' a piece of music sounds.

The correlation matrix obtained is:



Here we notice that there is high correlation between spectral centroid and rolloff, which is also seen in their perceptual similarity.