Assignment - 2 & 3 Report

P Romaharshan
2020101069

Question - A:

1. Harmonics:
   c) What are the perceptual differences between all 3 versions:
   Ans)

   Harmonics: provide a rich and full sound, adding harmonic fulfillment and complexity to the tune.

   Odd Harmonics: Compared to the full harmonic rendition, this version seems more "Hollow".

   Even Harmonics: Compared to the odd harmonics iteration, this one has a cleaner tone but less warmth.

2. Virtual Pitch:
   d) What is the perceptual difference between all 3 melodies
   Ans)

   Harmonic melody: provides a rich and full-bodied sound experience.

   Absence of the Fundamental Frequency: The harmonic series can be used to produce the desired pitch, but the pitch will still be audible even if it sounds a little thinner or less grounded.

   The absence of the fundamental and second harmonics makes the tune even more heavenly by reducing its richness. The absence of these lower harmonics may have a slight impact on pitch perception.

Question - B:

Part - 1:

The code has been provided in the zip folder. The below table tabulates the estimate and the results of the estimated and the actual values:

Number	Track Name	Estimated Tempo	Actual Tempo
1	Micheal_Jackson	120	185.3017

2	Dream_Threater	120	97.5394
3	Mozart	100	140.8711
4	Queen	100	109.4126
5	Taylor_Swift	40	51.5102

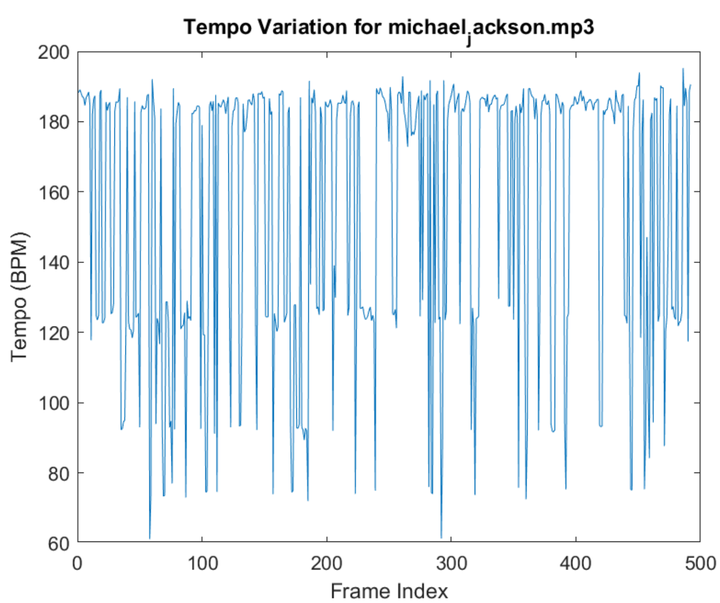
For tempo range estimation, as different excerpts have different tempos at different times, results are tabulated using same manner:

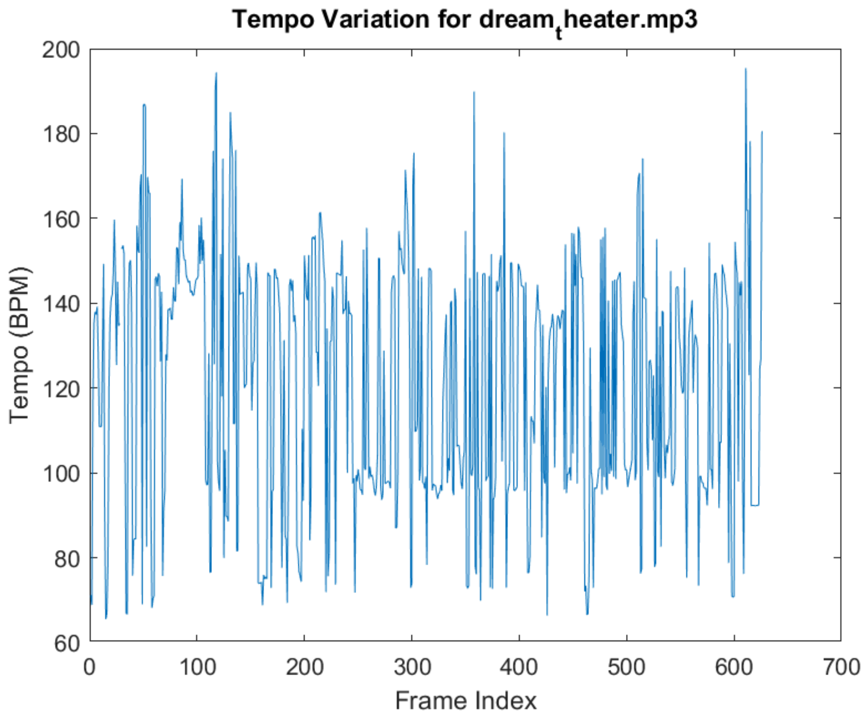
Number	Track Name	Min Tempo Estimate	Max Tempo Estimate	Min Tempo Actual	Max Tempo Actual
1	Micheal_Jackson	52	180	71.99	190.21
2	Dream_Threater	48	180	68.82	156.55
3	Mozart	48	220	61.21	194.22
4	Queen	40	160	61.09	196.62
5	Taylor_Swift	52	180	66.90	196.52

Note: To calculate the range of tempo we have used frame length of 2 and hop size of 0.25
The tempo plots obtained are in the zip file as well as attached below:

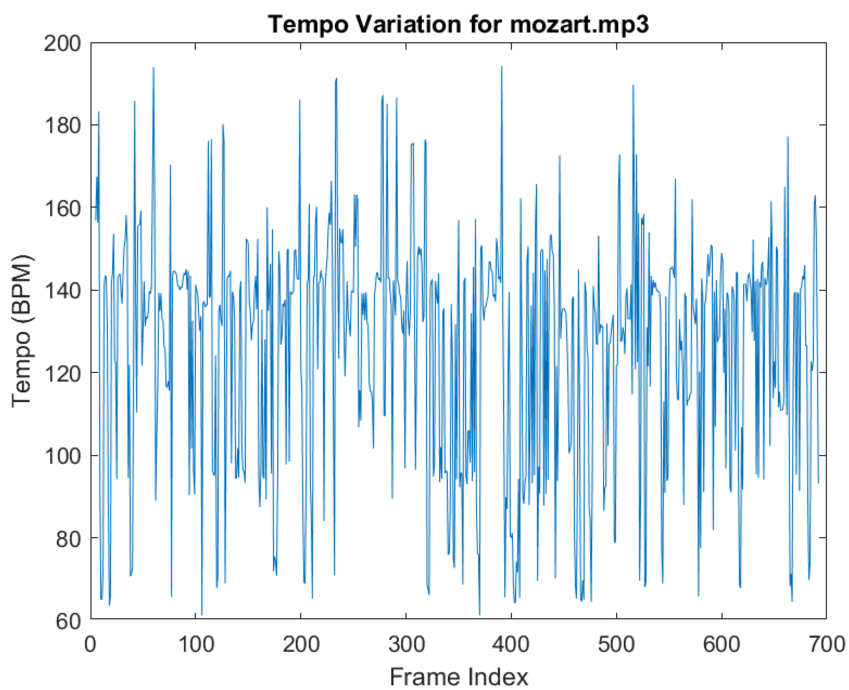
Part-2:

(Actual values are calculated using mirtempo function in MATLAB for the above table). The plots are given below:

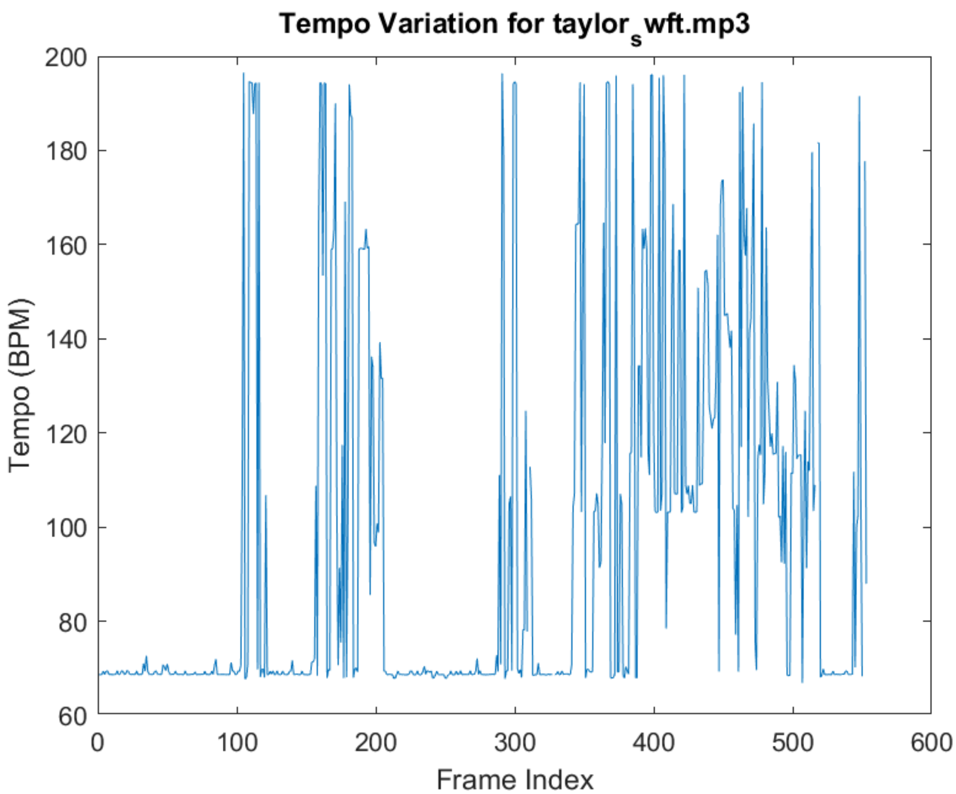


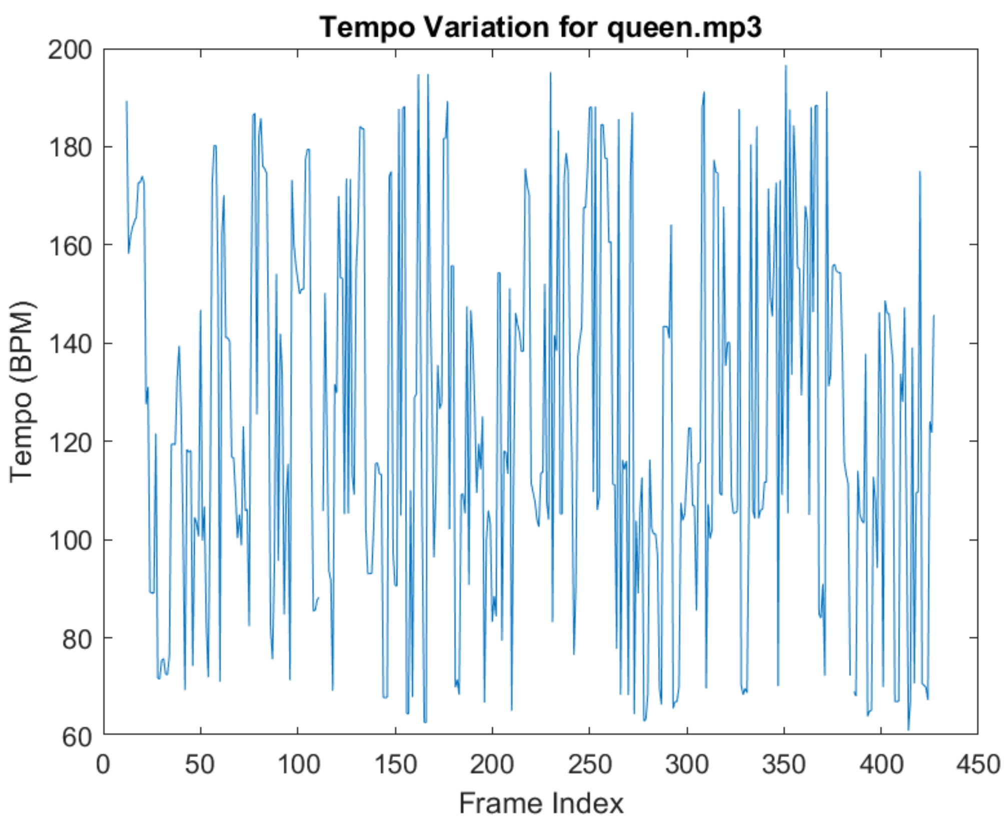


Tempo Variation for dream$_t$heater.mp3
Tempo (BPM)
Frame Index



Tempo Variation for mozart.mp3
Tempo (BPM)
Frame Index


Tempo Variation for taylor$_s$wft.mp3
Tempo (BPM)
Frame Index
200
180
160
140
120
100
80
60
0
100
200
300
400
500
600


Tempo Variation for queen.mp3
Tempo (BPM)
Frame Index
200
180
160
140
120
100
80
60
0
50
100
150
200
250
300
350
400
450

There is a general alignment between computational and perceptual estimates of tempo. The perceptual estimates are lower than the actual tempi but not too far. In these instances, computational assessments tend to offer a more precise reflection of the tempo.

Divergences in estimates can stem from a multitude of factors, including fluctuations in tempo within the excerpts, intricate rhythmic patterns, and variances in individual perceptual judgments. Furthermore, computational algorithms may not consistently capture subtle nuances, which can contribute to minor disparities between computational and perceptual assessments. The order of listening may also have an effect on the perceptual estimate.

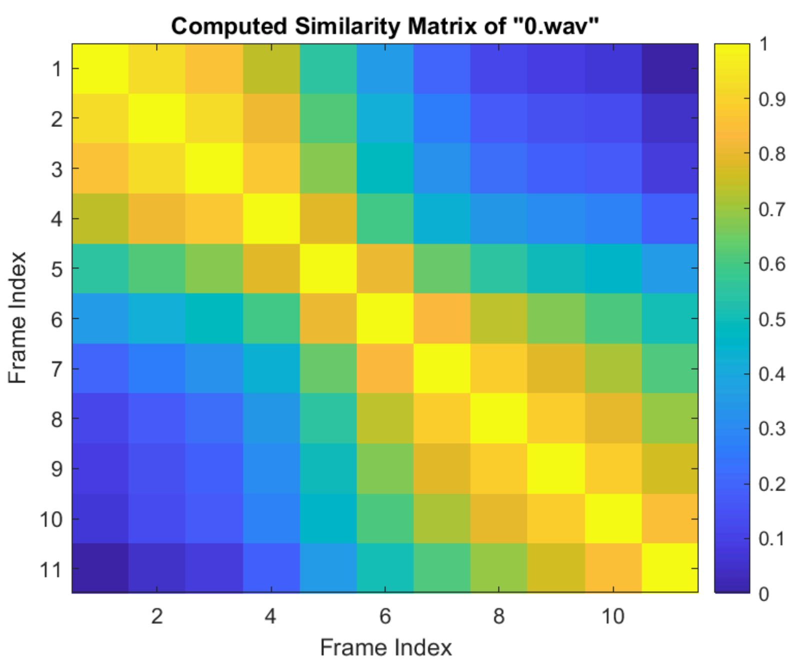
Part - 3:

Almost all of the tracks have variable tempi. The complex beats do not follow a definite tempo, though subtle to perception they are identified through computation. The above plots (in part 2 and Table 2) confirm it.

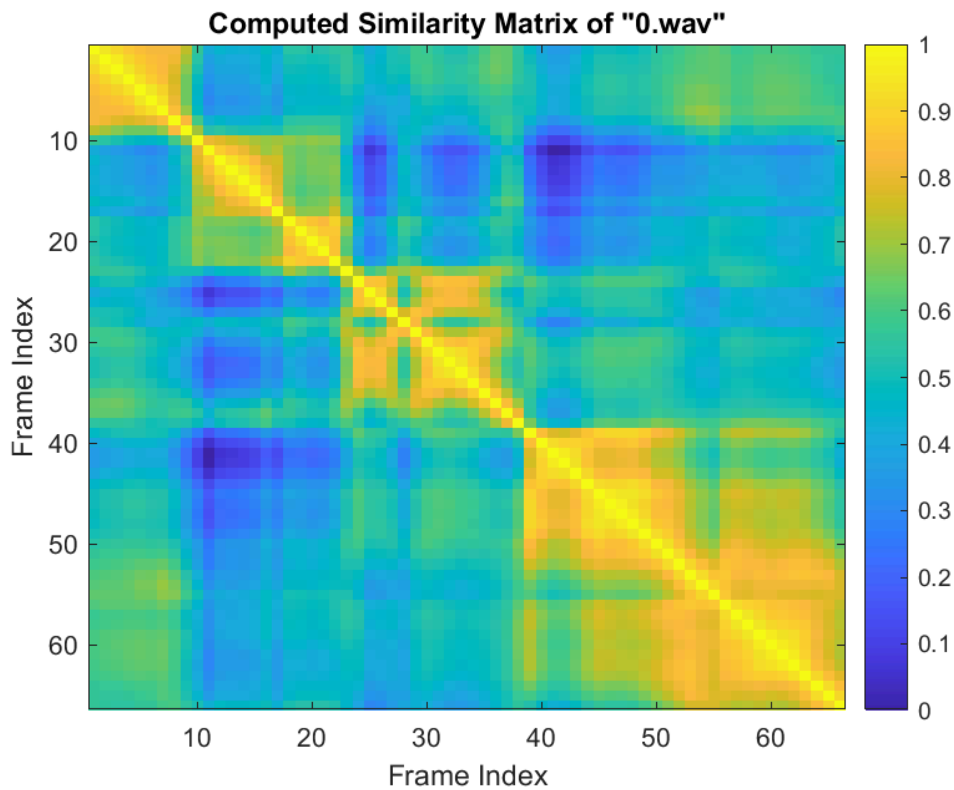
The reason is the same as the previous part.

Repetition in Music:

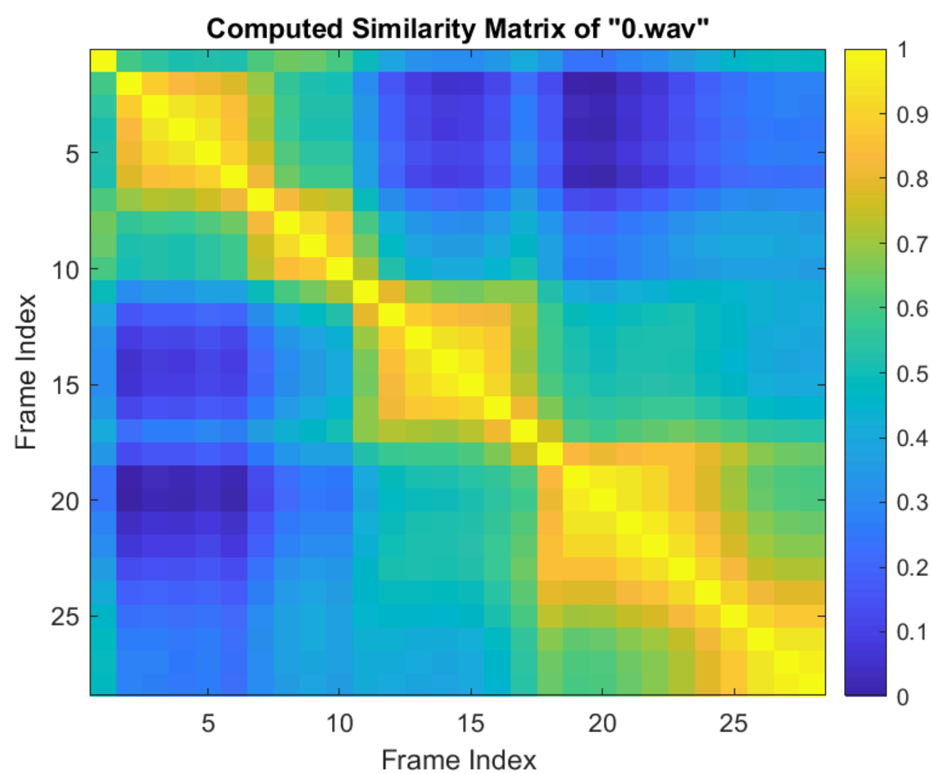
We analyzed the ‘0.wav’ file. We did chromagram analysis which gives the pitch distribution, frame decomposition using variable window length and hop size and obtained the following plots.



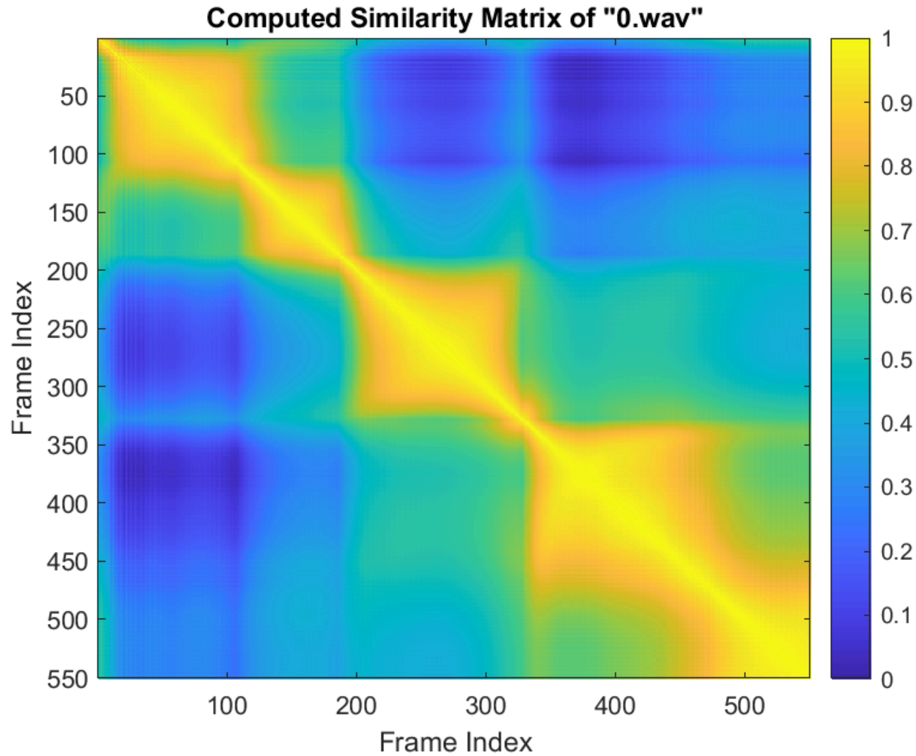
The window length is 1 and hop size if 0.05 (default).



The window length is 0.2 and hop size is 0.1



The window length is 0.4 and hop size is 0.1



The window length is 0.4 and hop size is 0.005

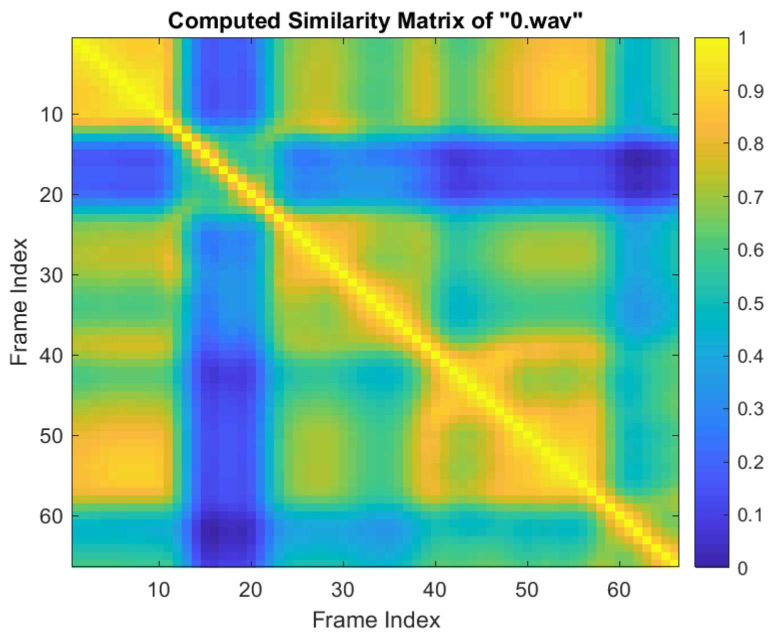
The effect observed was:

Chromagram Frames and Similarity Matrix:
Chromagram decomposition splits an audio signal into short-time segments called frames. Each frame is converted into a chromagram, which represents the presence of musical pitches across different frequency ranges. The similarity matrix is then constructed by calculating the similarity between these chromagram representations of different frames.

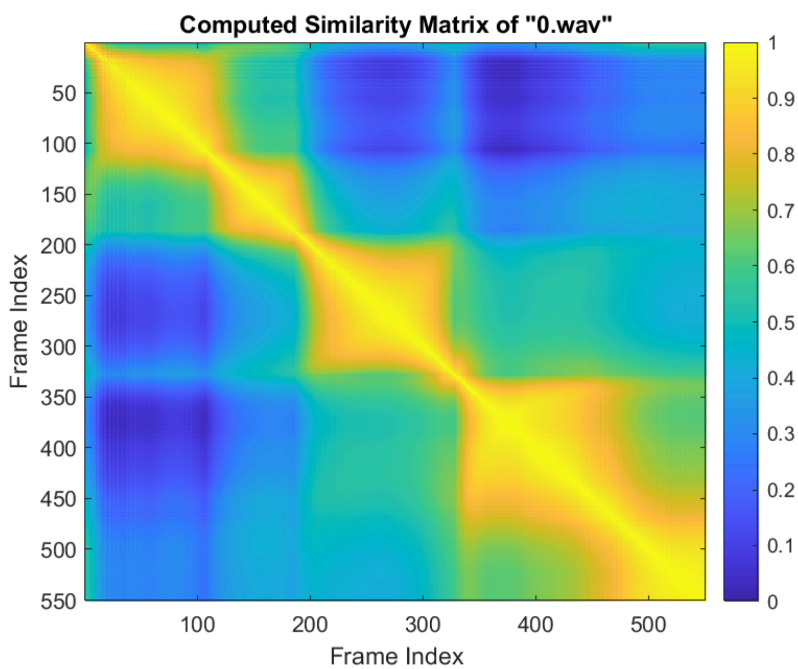
Window Length and Hop Size:
Window Length: This refers to the duration of each individual frame analyzed. A longer window length captures more detail from the audio but reduces the time resolution of the analysis. Conversely, a shorter window provides higher time resolution but less spectral detail.
Hop Size: This determines the amount of time between the center points of consecutive windows. A smaller hop size results in more overlap between frames, capturing more detail about how the music changes over time. A larger hop size reduces the overlap and provides a coarser analysis.

Impact on Similarity Matrix:
Window Length: A longer window length might lead to higher similarity scores between

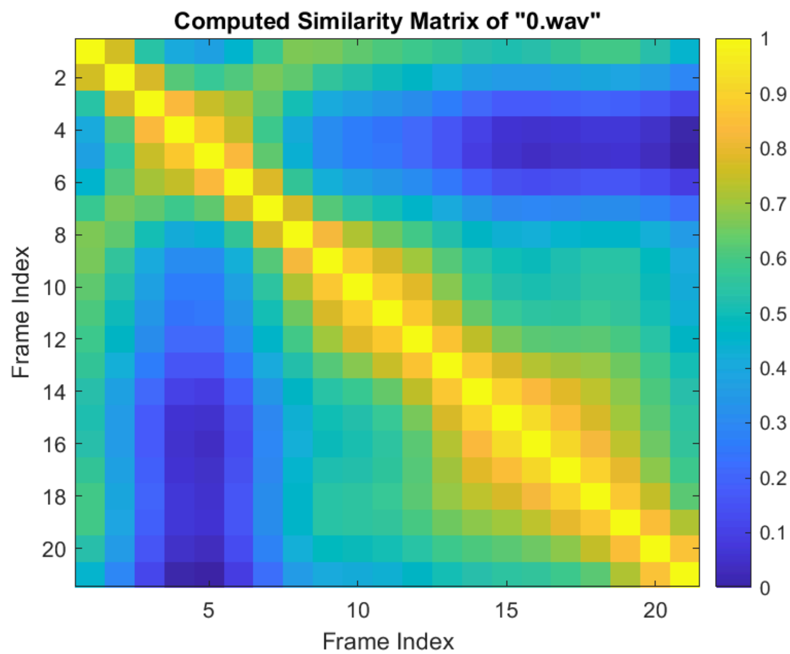
nearby frames because they will capture a larger portion of similar-sounding music. However, it might miss similarities between distant frames with shorter transitions.
Hop Size: A smaller hop size creates a denser similarity matrix with more data points. This can be useful for capturing subtle changes in the music. However, a larger hop size might be sufficient to identify broader sections with similar characteristics.
Changing the feature: We used spectrum (mirspectrum). The output of mirspectrum is a representation of how the energy of the signal is distributed across different frequencies.



Similarity matrix with a window length of 0.2 and hop_size of 0.1.



Similarity matrix with a window length of 0.4 and hop_size of 0.05.



Similarity matrix with a window length of 0.5 and hop_size of 0.1

How does Window Length and Hop Size affect the Similarity Matrix in this Case?

<u>Window Length</u>:
Longer window lengths provide better frequency resolution, meaning you can distinguish between different frequency components more accurately. However, they result in poorer time resolution, as you're averaging over a longer time span.

With a longer window length, the extracted features are more sensitive to changes in frequency content over time.

When computing the similarity matrix, a longer window length may lead to a more detailed analysis of similarities between segments of the signals in the frequency domain. It can capture finer spectral details but may miss rapid changes in the audio.

<u>Hop Size</u>:
The hop size determines how much the window shifts in time between successive frames. A smaller hop size means more overlap between consecutive windows.

Smaller hop sizes provide better time resolution, as you're capturing more frequent snapshots of the signal over time. However, they lead to increased computational complexity.

With a smaller hop size, the extracted features can capture rapid changes in the signal, providing a more detailed representation of temporal dynamics.

In the context of computing the similarity matrix, a smaller hop size may result in a higher density of comparisons between segments of the signals, potentially capturing more finer similarities.

Which Audio Feature is best suited for perceptual segmentation and repetition?

The choice of parameters in music segmentation using a similarity matrix significantly affects the final results. Here's how changing different aspects can impact the segmentation:

<u>Audio Feature</u>:
<u>Chromagram</u>: As discussed earlier, chromagrams capture the overall pitch content (pitch class) and are good at identifying harmonic changes. This makes them suitable for segmenting based on chord progressions and tonal shifts.

<u>Mel-frequency cepstral coefficients (MFCCs)</u>: MFCCs represent the spectral envelope of the audio, capturing how the energy is distributed across different frequency bands. They are good at capturing changes in timbre and instrumentation, which can be helpful for identifying sections with different musical textures.

<u>Spectrum</u>: The raw spectrum provides detailed information about all frequencies present at a given time. While it holds the most information, it can be computationally expensive to analyze and might lead to a very dense similarity matrix with less interpretable patterns.

<u>Impact</u>:
<u>Chromagram</u>: Good for segmenting based on harmony and tonal changes, might miss segmentation based on timbre or instrumentation changes.

<u>MFCCs</u>: Good for capturing changes in timbre and instrumentation, might be less sensitive to subtle harmonic changes.

<u>Spectrum</u>: Provides the most detail but can be computationally expensive and lead to noisy results.

<u>Frame Length and Hop Size</u>:
<u>Frame Length</u>: A longer frame captures more detail within a segment but reduces the time resolution. This can lead to merging short transitions or missing rapid changes. Conversely, shorter frames provide higher time resolution but might miss longer-term harmonic progressions.
<u>Hop Size</u>: A smaller hop size creates a denser similarity matrix with more data points, helping capture subtle changes. However, a very small hop size might lead to over-segmentation, while a larger hop size might miss important transitions.

Impact:
Longer frame length: Might miss short transitions or rapid changes, good for capturing longer harmonic progressions.

Shorter frame length: Good for capturing subtle changes, might merge short transitions.

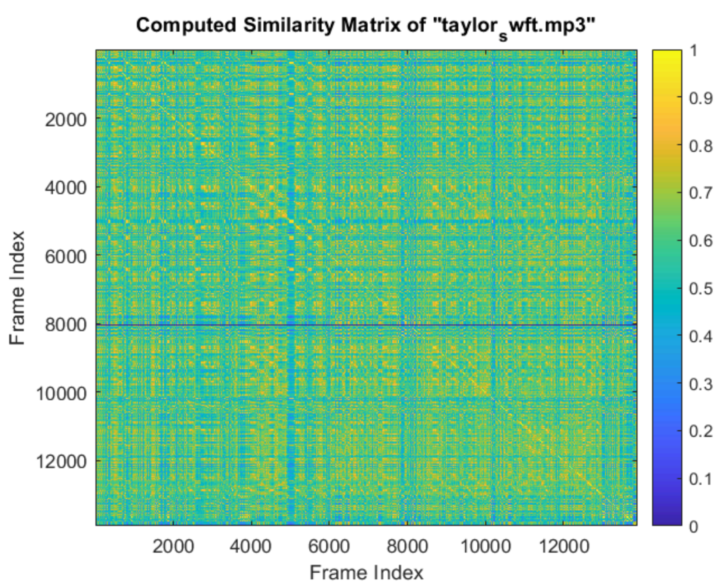
Smaller hop size: Denser matrix, captures subtle changes, might over-segment.

Larger hop size: Less dense matrix, might miss important transitions.

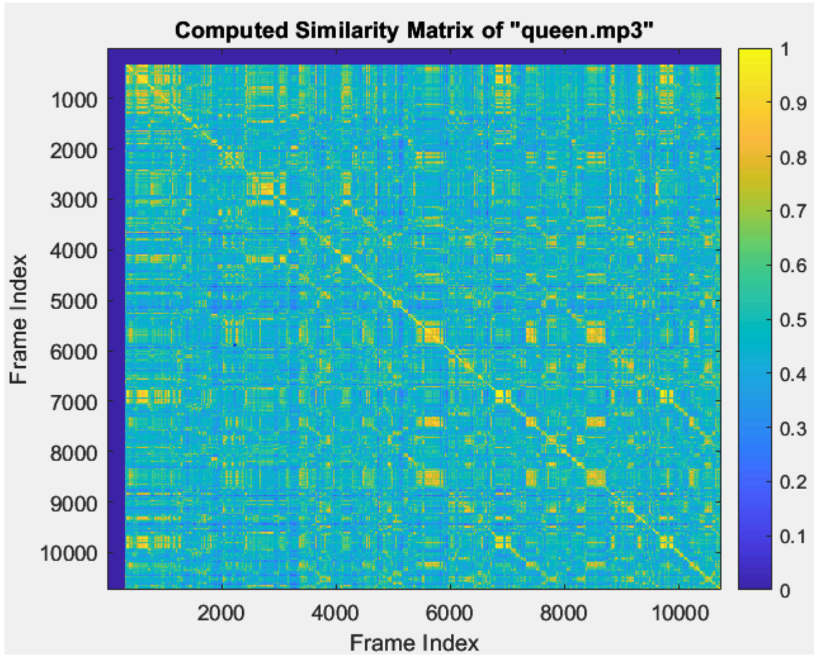
Best Feature for Perceptual Segmentation and Repetition:
The "best" feature depends on what aspect of perceptual segmentation and repetition you're most interested in:

Harmony and Chords: Chromagram is a strong choice as it directly captures pitch class information.
Timbre and Instrumentation: MFCCs are a good option as they effectively represent the spectral envelope.
Overall Similarity: Spectrum can be used, but it requires careful consideration of computational cost and managing the complexity of the similarity matrix.
Combining Features: In some cases, combining multiple features, like chromagram and MFCCs, can provide a more comprehensive picture by capturing both harmonic and timbral changes.
Note: *All the pictures of the analysis have been provided in the zip file. For the note, all analysis on the samples provided agree with the above drawn conclusions.*

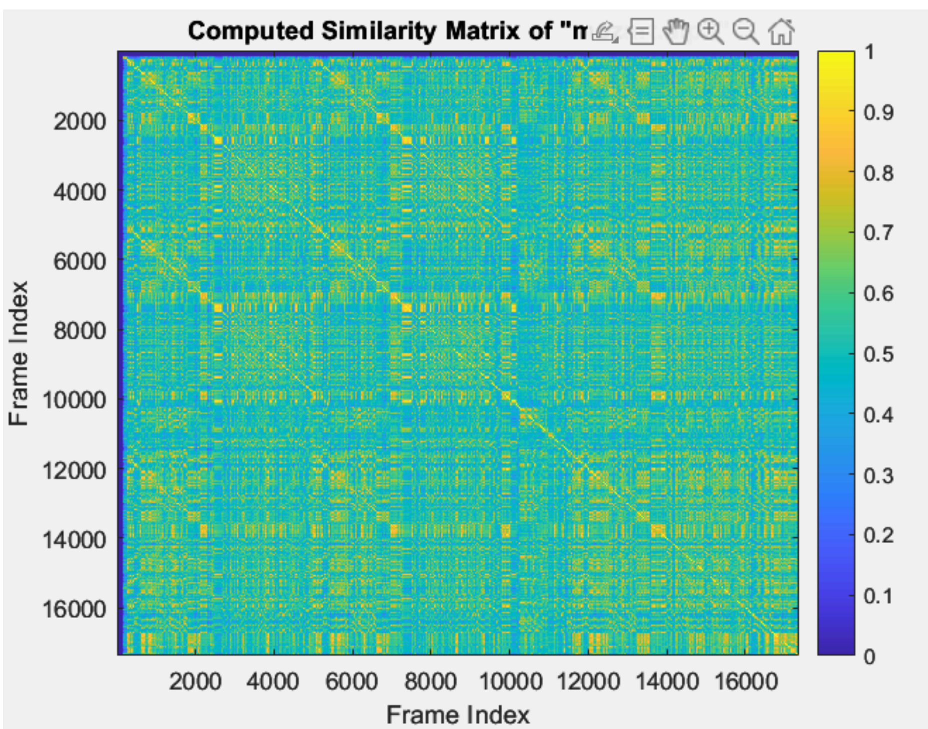
Some Similarity Matrix based of Chromagrams:



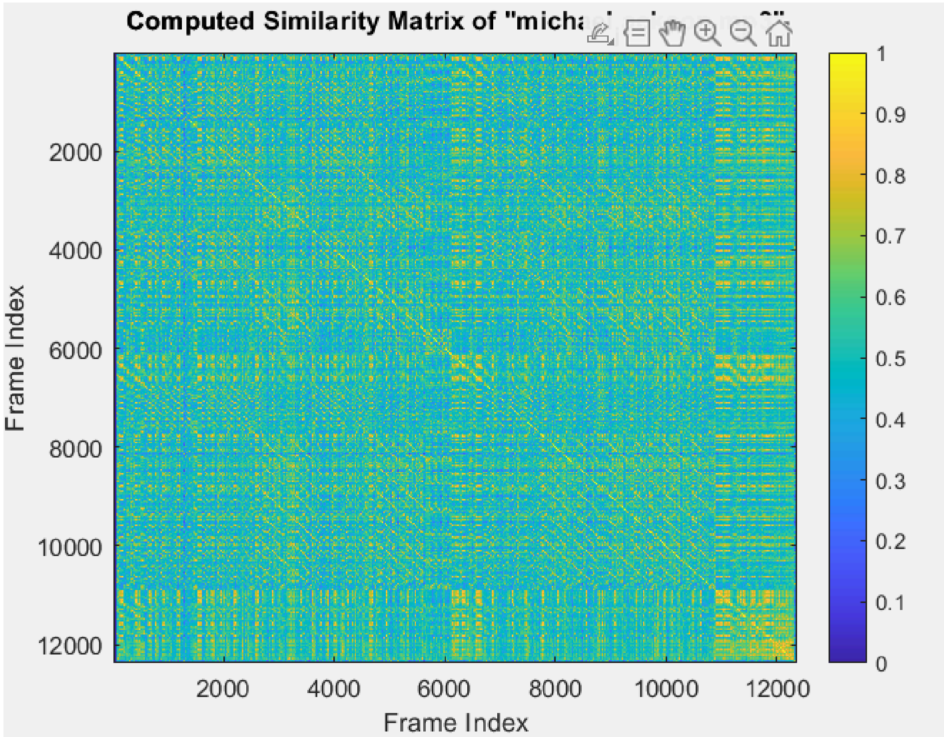
For the track taylor_swift, with a window size of 0.2 and hop_size 0.1



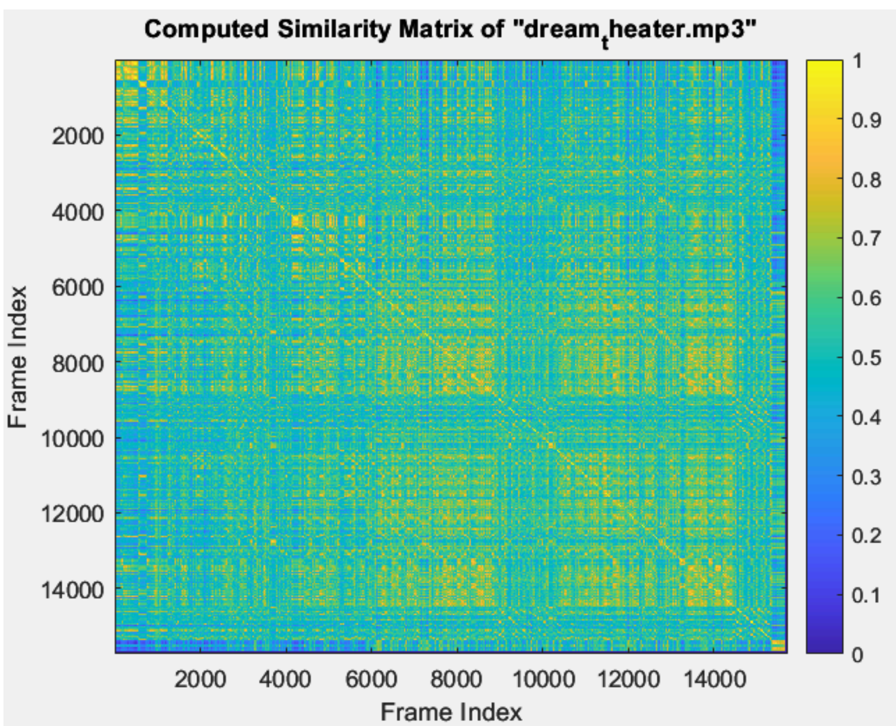
For the track queen, with a window size of 0.2 and hop_size 0.1



For the track mozart, with a window size of 0.2 and hop_size 0.1

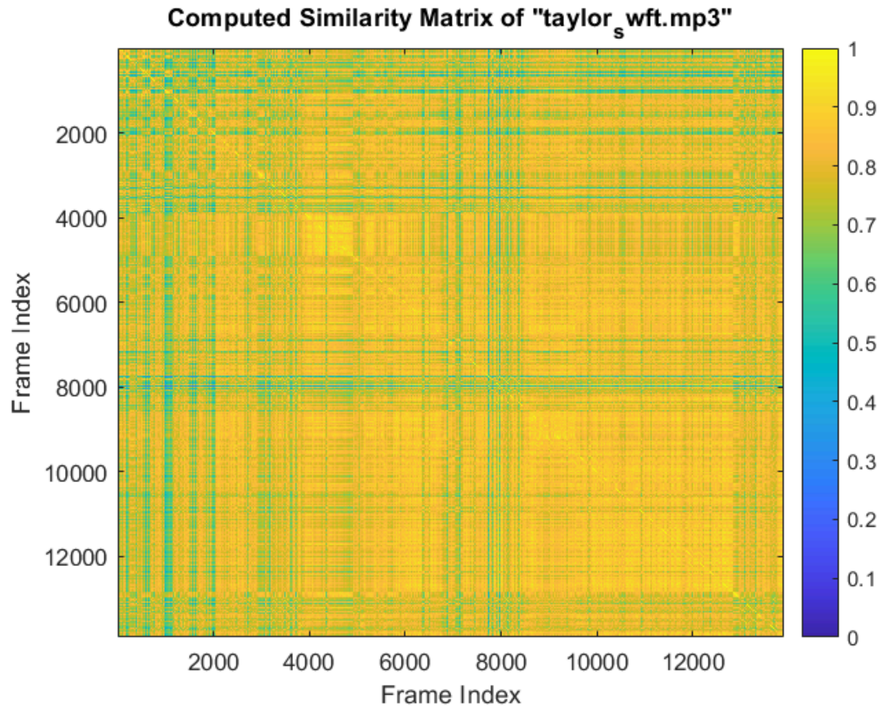


For the track micheal_jackson, with a window size of 0.2 and hop_size 0.1

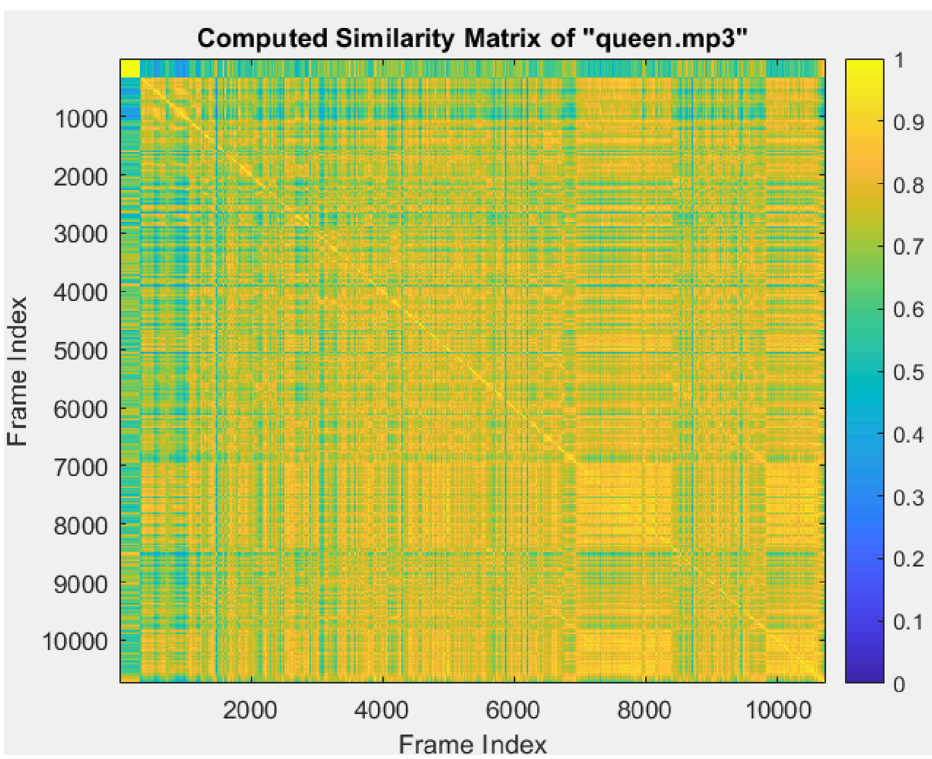


For the track dream_theater, with a window size of 0.2 and hop_size 0.1

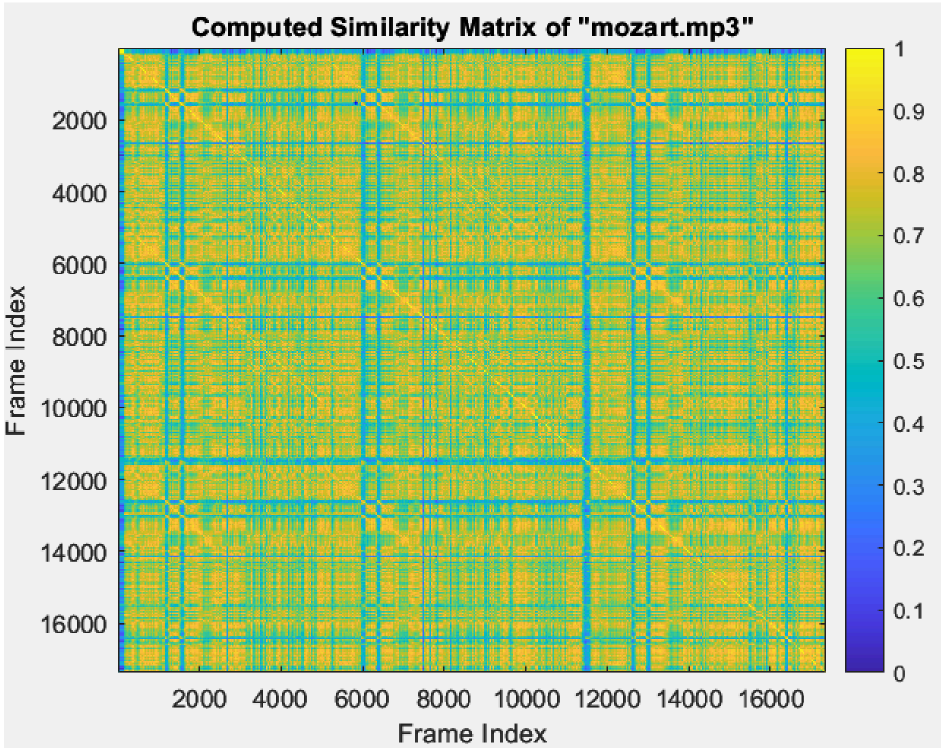
Some Similarity Matrix based of MFCCs:



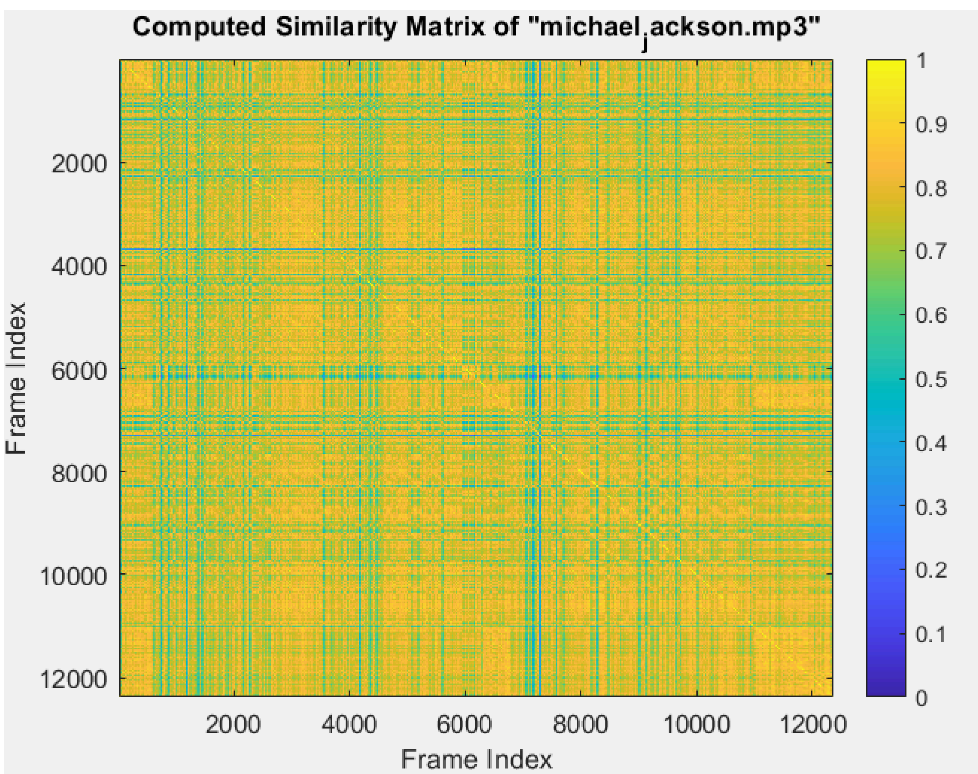
For the track taylor_swift, with a window size of 0.2 and hop_size 0.1



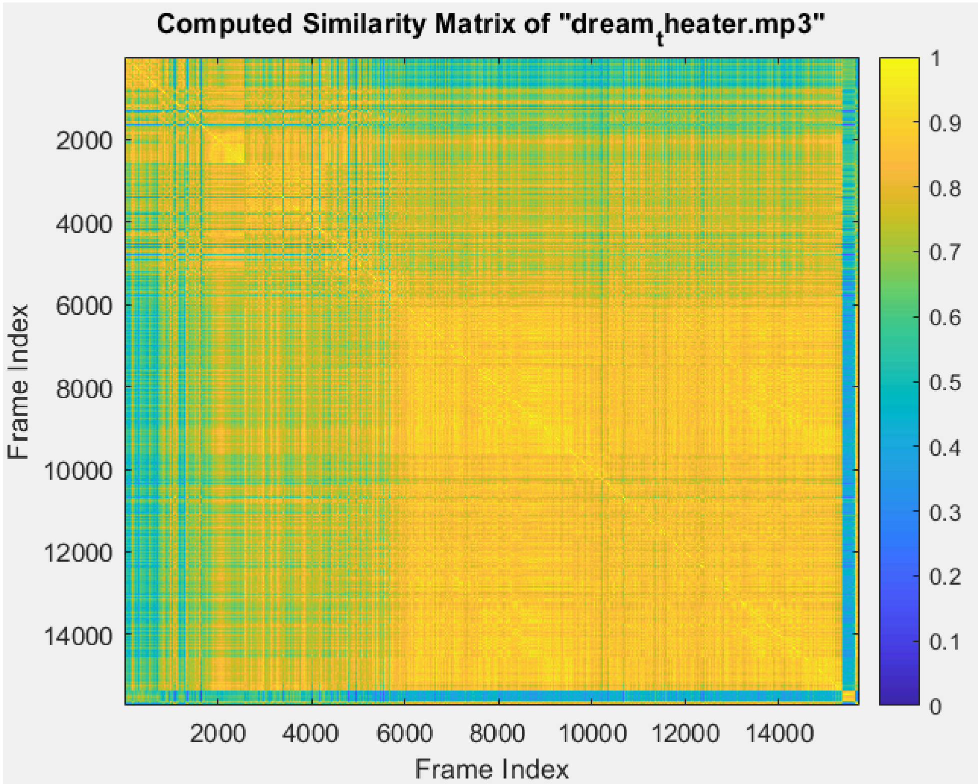
For the track queen, with a window size of 0.2 and hop_size 0.1



For the track mozart, with a window size of 0.2 and hop_size 0.1

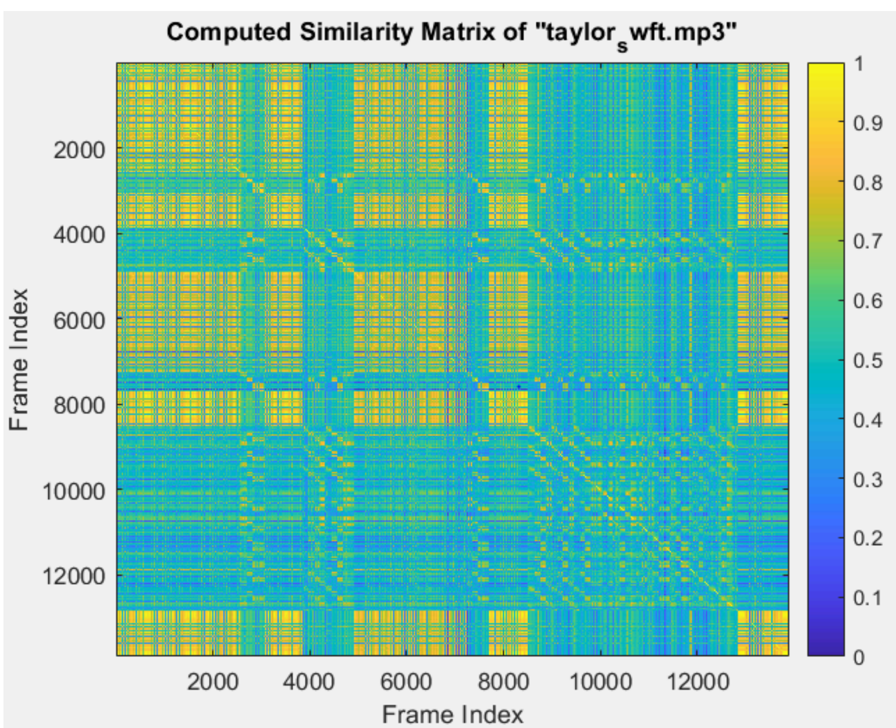


For the track micheal_jackson, with a window size of 0.2 and hop_size 0.1

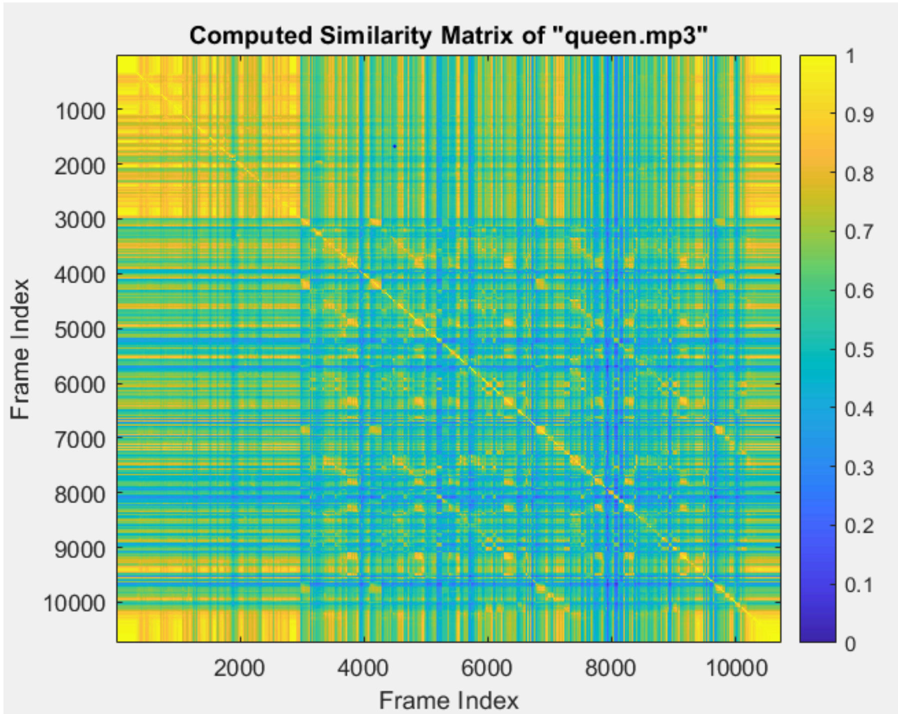


For the track dream_theater, with a window size of 0.2 and hop_size 0.1

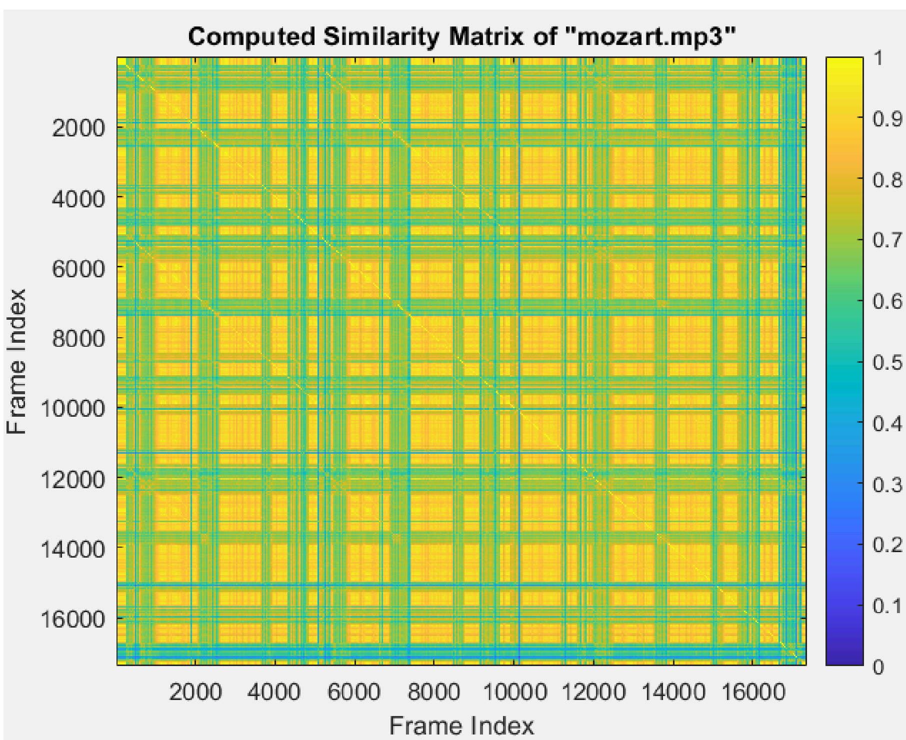
Some Similarity Matrix based of Spectrum:



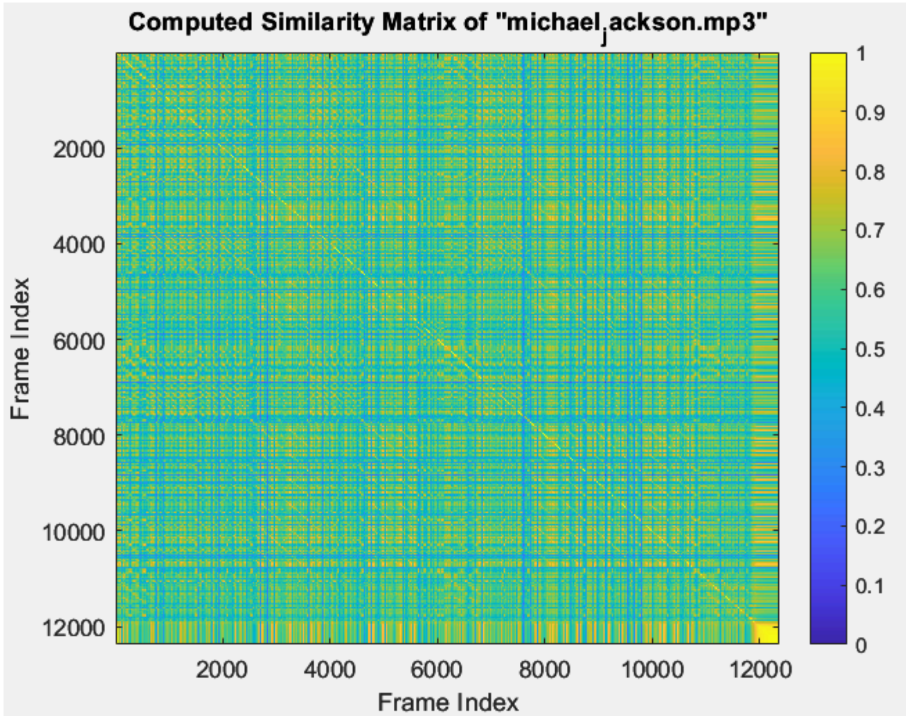
For the track taylor_swift, with a window size of 0.2 and hop_size 0.1



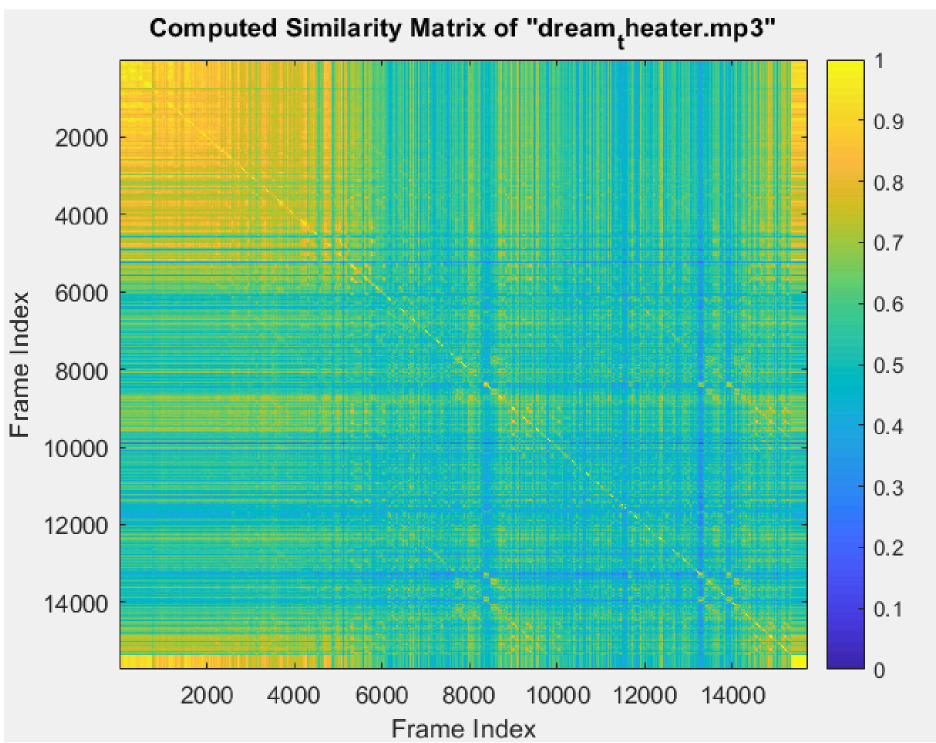
For the track queen, with a window size of 0.2 and hop_size 0.1



For the track mozart, with a window size of 0.2 and hop_size 0.1



For the track micheal_jackson, with a window size of 0.2 and hop_size 0.1



For the track dream_theater, with a window size of 0.2 and hop_size 0.1

Code: https://github.com/Romaharshan/MMT_Assignment_2-3/tree/main/codes
Result: https://github.com/Romaharshan/MMT_Assignment_2-3/tree/main/results