A Preliminary Analysis of CRIM Centroids based on Harmonic CRIM Interval Slicing

Edgar Leon

Department of Music, Haverford College

MUSC H255: Encoding Music

Professor Freedman

Introduction

CRIM (Citations: The Renaissance Imitation Mass) has allowed different kinds of analysis of Renaissance music dealing with note distribution, cadence types, cadence analysis, harmonic intervals, and more. In doing harmonic trace analysis of different masses, the idea of a "centroid" or summary of intervals that represent a piece has come about. The summary of Bass-Tenor, Bass-Alto, and Bass-Soprano intervals when plotted in 3D space, allows for the visualization of distance between different musical works, where a closer distance indicates two pieces are more similar. How well a piece is represented by these centroids is unknown, in fact, it is unknown if the centroids even represent the piece to any degree of accuracy or if they are just extraneous data. Although some aspects of CRIM have been studied in detail, the idea of centroids is relatively new and the significance of them deserve to be explored further. This project set out to answer the following question: What data can we extrapolate from CRIM centroids and is it significant enough to continue to use them in further data analysis? The findings from this paper suggest that CRIM centroids do in fact have some significance and do represent a piece, composer, and movement to an extent.

Results

The data suggests that CRIM centroids do in fact have statistical significance to a degree, even though they are only a "summary" of the entire Mass. Without running any regressions and simply looking at the 3D scatterplots, both grouped by movement and composer, it is evident that there is a positive correlation in the data.

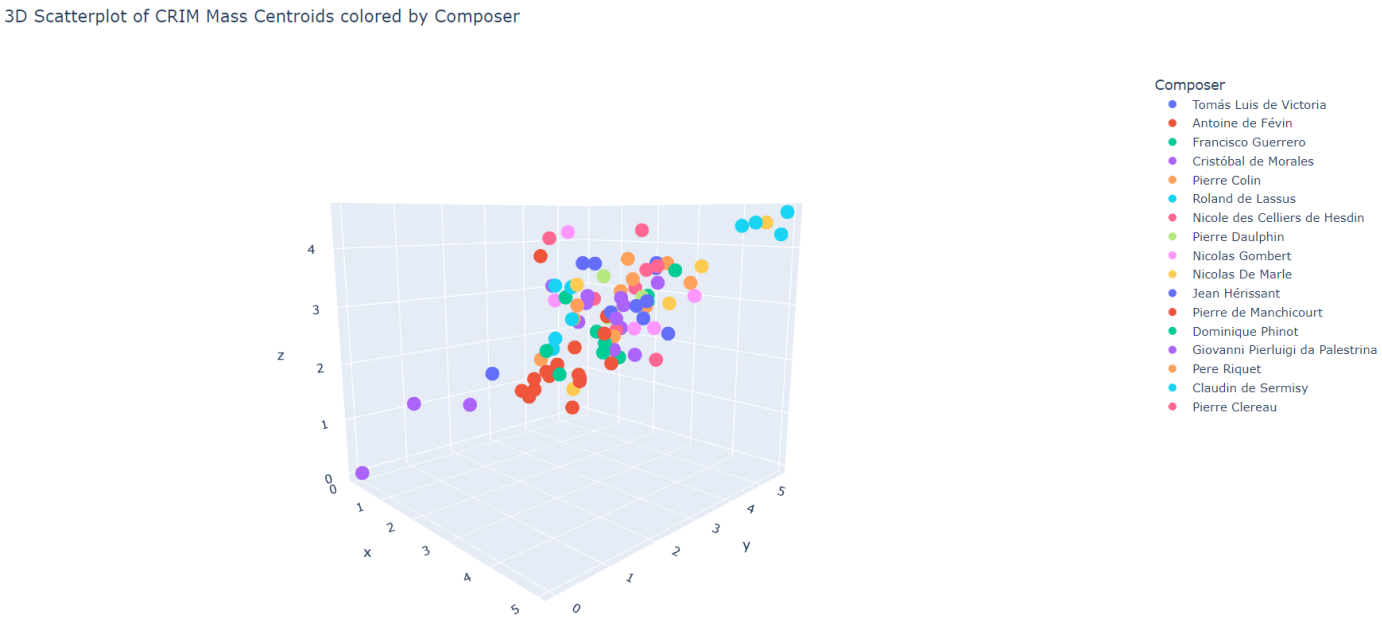

Fig 1. A 3D Scatterplot of CRIM Mass Centroids colored by Composer where x represents the Bass-Tenor interval, y represents the Bass-Alto interval, and z represents the Bass-Soprano interval.

The graph illustrates that each composer seems to have a specific cluster of space that they occupy, and certain composers tend to cluster towards the center, whereas others tend to cluster towards the edges, thus making them outliers. For example, it seems as if Palestrina is clustered in the center, whereas Lassus is clustered in the top right.

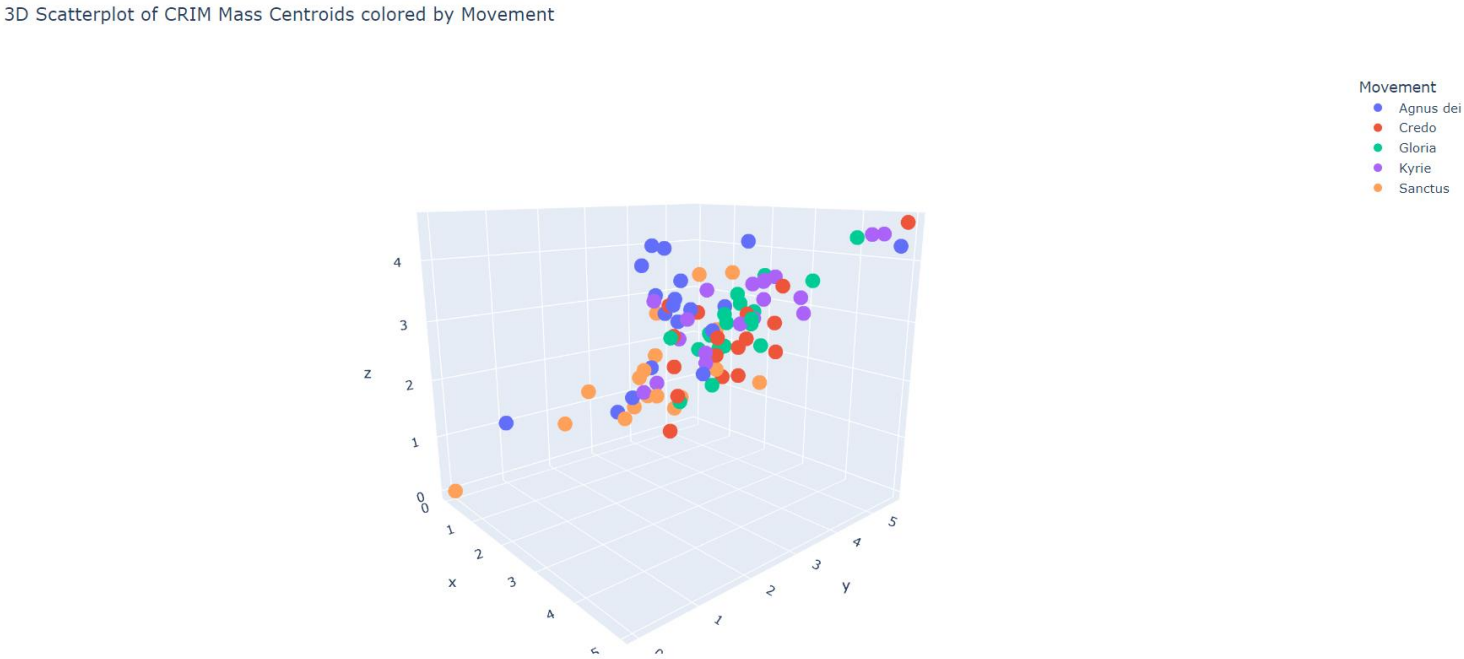

Fig 2. A 3D Scatterplot of CRIM Mass Centroids colored by Movement where x represents the Bass-Tenor interval, y represents the Bass-Alto interval, and z represents the Bass-Soprano interval.

The graph illustrates that each movement also seems to have a specific cluster of space that they occupy, and certain movements tend to have wider spread of data points, whereas others tend to have less of a spread. For example, it seems as if the Sanctus pieces are clustered in the center, but the spread is in the lower half of the graph, whereas the Kyrie pieces are also clustered in the center, but the spread is in the upper half of the graph.

After running a linear regression and graphing the correlation plots, there is evidence to suggest that not only is there a positive correlation between the CRIM centroid coordinates and Composer of the Mass, but there is also one composer that is far different from the rest.

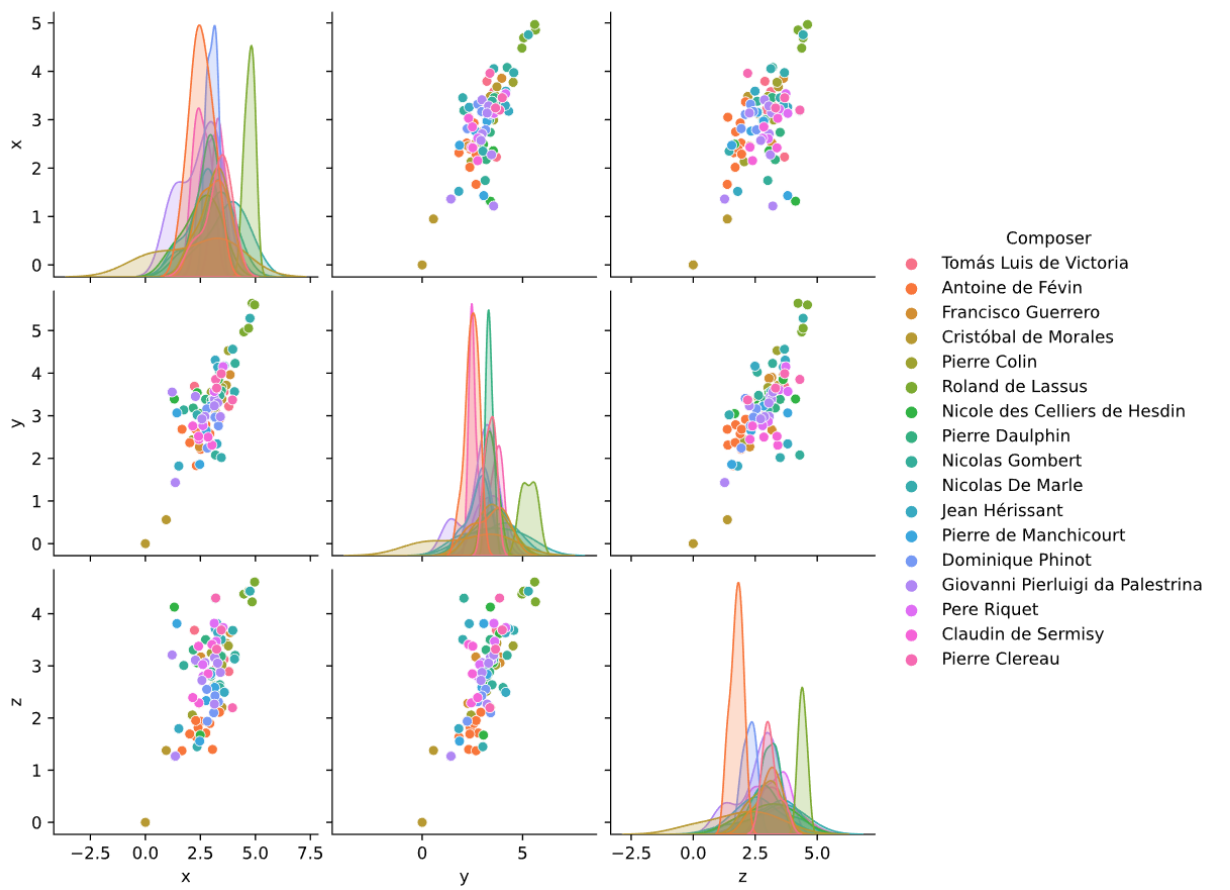

Fig 3. A Correlation plot of CRIM Mass Centroid Coordinates and Composers where x represents the Bass-Tenor interval, y represents the Bass-Alto interval, and z represents the Bass-Soprano interval.

The correlation plot shows a positive correlation with all coordinates, but it most notably shows a strong positive correlation between the Bass-Tenor and Bass-Alto interval. An interesting observation, however, shows that one composer has a significantly different distribution of data that the others, Roland de Lassus.

In order to verify the statistical accuracy of this data, one could fit the data with a plane of best fit and measure the distance from the points to the plane. The further the points are from the plane of best fit, the less predictable they are by the model. Because the plane of best fit causes the data points to be compressed on the graph to the point where it is incomprehensible, viewing a table of the average distance from the plane grouped by composer will yield better data interpretation.

```
Composer
Roland de Lassus                      6.185050
Nicolas De Marle                      4.883572
Pierre Clereau                        4.632764
Tomás Luis de Victoria                4.371431
Pierre Colin                          4.302930
Nicolas Gombert                       4.276606
Francisco Guerrero                    4.217173
Dominique Phinot                      4.146737
Pere Riquet                           4.119849
Jean Hérissant                        4.096760
Pierre Daulphin                       3.997574
Nicole des Celliers de Hesdin         3.716111
Antoine de Févin                      3.607814
Pierre de Manchicourt                 3.593615
Giovanni Pierluigi da Palestrina      3.528974
Claudin de Sermisy                    3.500795
Cristóbal de Morales                  3.236383
Name: Distance_From_Plane_by_Composer, dtype: float64
```

Fig 4. A Table Output of the Average Distance from the Plane of Best Fit of Composer Based on CRIM Mass Centroid Coordinates

Distances	Q1	Q3	IQR	Lower Bound	Upper Bound
6.18505	3.23638	4.33718	1.1008	1.58518675	5.98837675
4.88357					
4.63276					
4.37143					
4.30293					
4.27661					
4.21717					
4.14674					
4.11985					
4.09676					
3.99757					
3.71611					
3.60781					
3.59362					
3.52897					
3.5008					
3.23638					

Fig 5. A Table Output of the Average Distance from the Plane of Best Fit of Composer Based on CRIM Mass Centroid Coordinates with the Upper Bound representing outliers

The distance from the plane of best fit by Composer shows that Roland de Lassus is furthest from the plane and Cristobal de Morales is closest. In other words, Lassus is the least predictable or the one that differs the most from the other composers; the only outlier (see Fig 5), and Morales is most accurately represented by the plane of best fit.

Applying the same concept to the movement types, yields the following results.

```
Movement
Sanctus      1.236972
Credo        1.663816
Gloria       2.002655
Agnus dei    2.145261
Kyrie        2.237014
Name: Distance_From_Plane_by_Movement, dtype: float64
```

Fig 6. A Table Output of the Average Distance from the Plane of Best Fit of Movement Type Based on CRIM Mass Centroid Coordinates

The distance from the plane of best fit by movement type shows that the Sanctus pieces are more accurately predicted by CRIM Mass Centroid Coordinates, whereas the Kyrie pieces are the ones that are less accurately predicted.

Conclusion

I initially planned on trying to create a network of similarity between the different composers, the masses, models, etc., but switched my idea after talking with Fabian about CRIM centroids, 3D visualization, etc. Before I talked with Fabian, I was trying to visualize all CRIM pieces even ones where there were more than 4 voices. After discussing with Fabian, I decided to try to run harmonic interval slicing on only masses with four voices and attempt to visualize them to see if there was any noticeable correlation between them. It was during this conversation that he mentioned that it is currently unknown if CRIM centroids even held any data significance at all…centroids were just data points that one could extrapolate from the multitude of intervals. After plotting the centroids and seeing that there seemed to be a trend in the data, I decided to look at the centroids grouped by composer and movement type and see what I could find. The correlation plot then revealed that the CRIM centroids coordinates had a positive correlation in predicting the composer. In other words, each composer does have a distinct style of composing. However, there was one composer that stood out, Roland de Lassus. I tried to run a SVM 3D Regression Surface Model on the data, but after talking with Fabian, we both agreed that it was not going to tell me very much about the data, so I decided to go with a plane of best fit or a tangent plane approach. In doing so, I was able to plot a normalized plane for all of the centroids and then calculate the distance from each point to that plane to try to see with composers and pieces were furthest from the plane. This data revealed that Lassus was yet again an outlier when it came to the plane of best fit and he was far different from the others, which supported the correlation plot findings. It seems like CRIM centroids do represent the pieces to some extent since it is possible to predict the composer and movement type, however, it is unknown how much more accurate all of the intervals can be when ran through the same data wrangling and analysis. Throughout this process, I familiarized myself more with the Plotly library, 3D

graphing and visualization, Linear Regressions, Statistics, and more. Compared to my goals at the start of the semester, I never thought I'd be capable of this level of data visualization and analysis.