Lab Session 6:
Audio analysis and feature extraction

Audio processing, video processing and computer vision

uc3m | Universidad Carlos III de Madrid

Made by:
Marina Gómez Rey (100472836)
María Ángeles Magro Garrote (100472867)

INDEX

1. Abstract

In this lab, we will explore various audio features for analyzing audio signals, specifically focusing on the classification of major and minor chords. By calculating key descriptors and understanding their relevance to this task, we aim to evaluate how each feature contributes to distinguishing between the two chord types.

2. Feature Extraction

1. Zero Crossing

Zero crossing rate definition

Zero crossing rate (ZCR) will measure how often a signal crosses the zero axis. Due to this, it will be a good feature in order to classify among noisy and smooth signals. A comparison among major and minor chords can be done in terms of ZCR in order to see if there is a significant difference among both.

Is zero crossing a key feature for this problem?

When comparing the amplitude vs. time plots *[Image 1, Image 2]* of major and minor signals, **no clear distinction** can be observed between them, as both show numerous crossings along the x-axis. Additionally, the Kernel Density Estimation (KDE) plot *[Image 3]*, which estimates the probability density function (pdf) of a random variable, reveals that the **distributions of both major and minor signals are quite similar**. This suggests that their Zero Crossing Rates (ZCR) are distributed in a comparable manner, pointing to similar underlying patterns and characteristics for both signal categories. Similarly, the boxplot *[Image 4]* shows almost **identical means and quartiles for both signals**. Consequently, since ZCR offers no additional information for classification, it is logical to exclude it from the analysis. In fact, removing it improves the AUC of the model.

2. Chroma STFT

Chrome STFT definition

The Chroma STFT captures the harmonic content of music by analyzing the energy distribution across the 12 pitch classes, helping to distinguish between major and minor chords. The values from the STFT represent the energy of specific frequencies at different times. In a **major chord, the STFT shows higher energy at the root, major third, and perfect fifth**, while a **minor chord displays a different energy distribution, particularly due to the minor third**. This difference in harmonic structure allows for differentiation between major and minor chords.

What is the output shape of this feature?

The output of Chroma STFT is a matrix with dimensions (num_frames, 12), where **num_frames depend on the frame length and hop size**, and 12 represents the 12 chroma bins for each pitch class.

Which frame length and hop should we choose?

The frame length refers to the size of each audio segment analyzed in Chroma STFT, affecting how much frequency detail can be captured. **A larger frame length improves frequency resolution** but reduces temporal precision, while a **smaller frame length enhances temporal resolution** but sacrifices frequency detail. The hop size controls how much the analysis window shifts between successive frames, affecting temporal resolution. A **larger hop size reduces temporal resolution**, while **a smaller hop size increases it**. Together, these parameters balance frequency and time resolution.

The classification results of different sizes can be tested comparing the AUC with our final model *[Image 5]*. As a result, frame length X and hop size Y are chosen. These results are reinforced by our posterior tuning. This combination happens to offer the best trade-off between capturing sufficient temporal information (with a moderate frame size) and the ability to process the data efficiently (with a smaller hop size).

3. Maximizing AUC

Studied features

1. Energy entropy - the plot of time vs amplitude can be used to compare the major and minor signals which in fact have a similar structure *[Image 1, Image 2]*, with a thicker part at the beginning that then becomes more plain. However, it seemed that the minor chord was in general a little bit thicker than the major. Although not being completely crucial, it improved our AUC.

2. Spectral entropy - represented by the spectrogram *[Image 6, Image 7]* and the entropy in the time frame *[Image 8, Image 9]* plot. In the first one, the main differences are located between frequencies 128 and 256. Also, the time graph shows more peaks for the major chords, while the spectrogram is more detailed in the minor chords, which makes sense because of the inverse relationship that exists between these two types of representations. Overall, this feature has proven to be very relevant, increasing the results in the AUC.

3. Mel-frequency Cepstral Coefficients (MFCCs) - Our hyperparameter search showed that using 13 time frames yielded the best results, capturing significant differences between major and minor chords [Image 10, Image 11]. These variations highlighted key characteristics of the chords, significantly improving the AUC and demonstrating their effectiveness in extracting meaningful patterns, that is why many statistics were obtained.

4. Spectral centroid - the study of this characteristic was made with the magnitude of the frequency graph *[Image 12, Image 13]* where it is clear that there is a difference in the frequencies for both types of chords: minor chords have more minima, maxima and variation than the major. Also, the ranges of the magnitudes are different, being bigger for the major chords. Again, it provided an improvement in the AUC.

5. Harmonic ratio - the harmonic ratio of each signal is calculated as the ratio of harmonic energy to total energy, representing the harmonic content of the signal. When we plot the density distributions for major and minor chords *[Image 14]*, we observe a slight difference between them, with minimal noise in the signals. Notably, the distribution reveals that major chords generally exhibit a higher harmonic ratio, indicating they are more harmonic. This subtle but consistent difference enhances the model’s ability to distinguish between the two types of chords, improving the AUC score.

6. Chroma features - represented by the chromagram of major and minor chords *[Image 15, Image 16] where* the differences between chords can be observed, which are very pronounced. Major chords, as expected, have more energy (more red) in general but also in the major pitch classes. On the contrary, minor chords have less energy (more blue) and the energy tends to concentrate on the minor pitch classes. This feature has proven to be very relevant, improving our results considerably.

7. Spectral spread - as visible in the frequency magnitude plot *[Image 12, Image 13]*, the spread of both types of chords is very similar. As a result, it introduces noise into the analysis, which negatively impacts the results. Therefore, this approach was not utilized.

8. Spectral flux - this measures changes in a signal’s power spectrum over time. High flux indicates rapid changes, while low flux suggests stability. The violin plot of major and minor chords *[Image 17]* can be studied to spot the differences in the classes. The AUC improves significantly with this feature.

Extra steps

Tuning the SVC model - the parameters of the model were tuned to optimize the AUC, and it resulted in a drastic enhancement of the AUC. The parameters involved in the tuning were C, gamma and the kernel, which remained as rbf, logical for a non-linear problem.

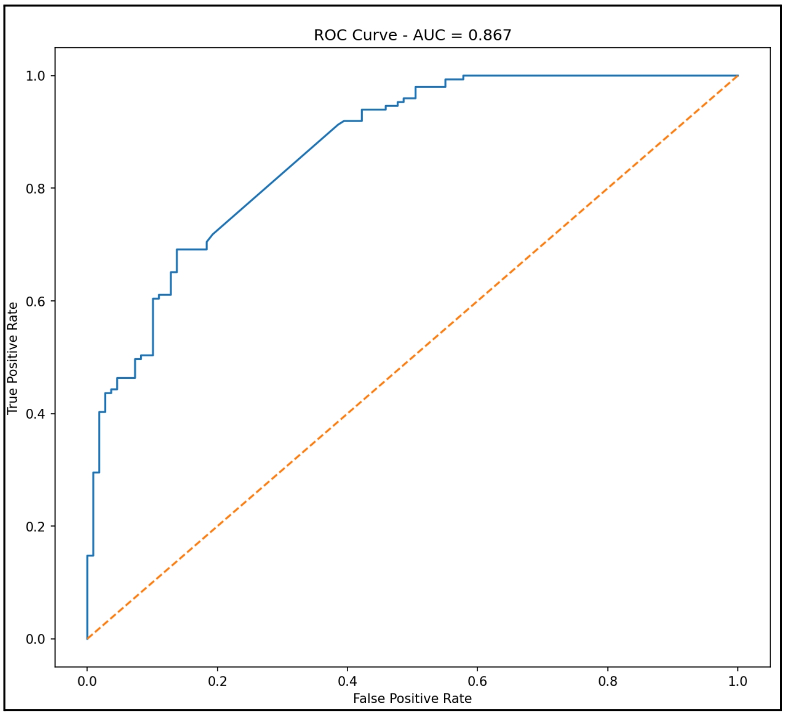
Erasing correlated values - our model was composed of a high number of variables, and the correlation among them should be studied in order to avoid noise production. To make feature selection we made the correlation matrix *[Image 18]* and decided to take out correlations higher than 0.8, which resulted in AUC improvement.

3. Graphic summary

SUMMARY TABLE

Experiment	AUC
1. Base template	**0.505**
2. 1 + Tuning	**0.522**
3. 2 + Zero Crossing Feature in	**0.499**
4. 2 + Plus Spectral entropy (no zero-crossing rate)	**0.526**
5. 4 + Spectral centroid	**0.543**
6. 5 + Spectral spread	**0.539**
7. 5 + Plus MFCCs (no spectral spread)	**0.572**
8. 7 + harmonic ratio	**0.575**
9. 8 + chroma features	**0.586**
10. 9 + spectral flux	**0.605**
11. Tuning of svc model (10)	**0.799**
12. Upgrading the MFCC statistics (11)	**0.861**
13. Taking out correlated variables (12)	**0.867**

ROC CURVE OF THE FINAL MODEL



IMAGES

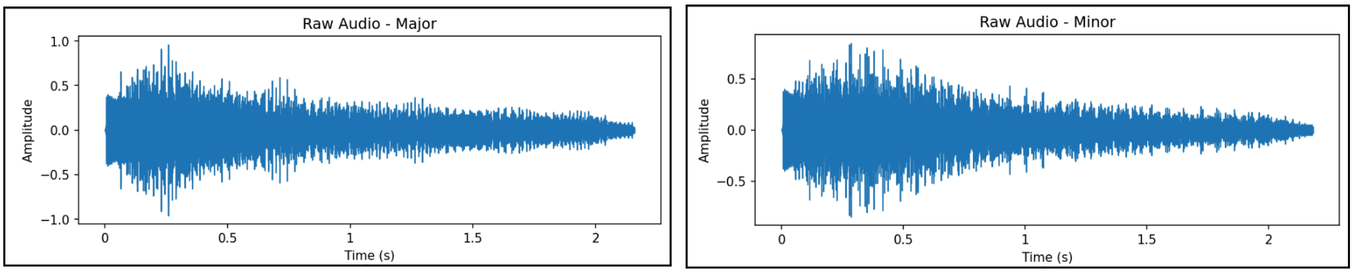


Image 1 (left) and 2 (right): amplitude vs time of a major and minor signal respectively

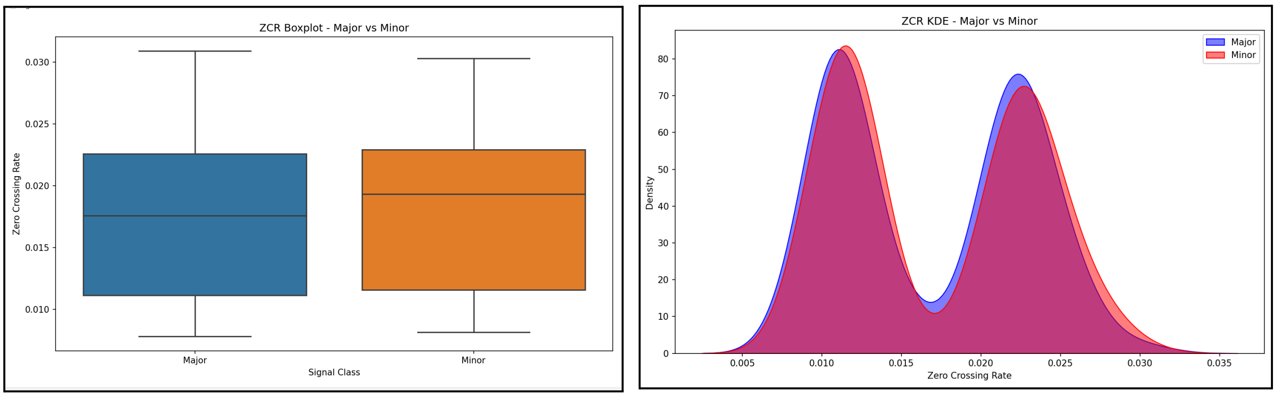


Image 3 (left) and 4 (right): boxplots and KDE plots for ZCR among minor and major signals

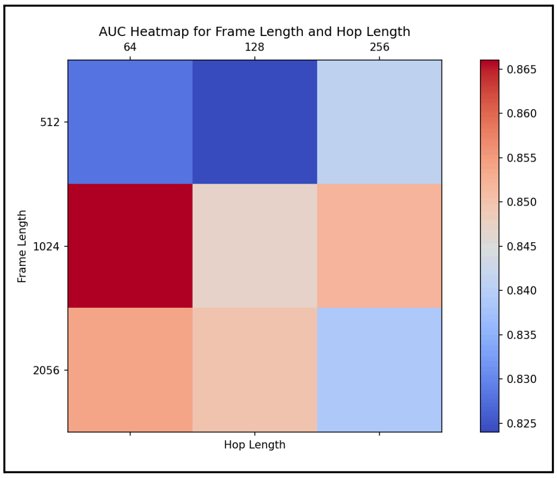


Image 5: AUC comparison for different frame lengths and hops in the final model

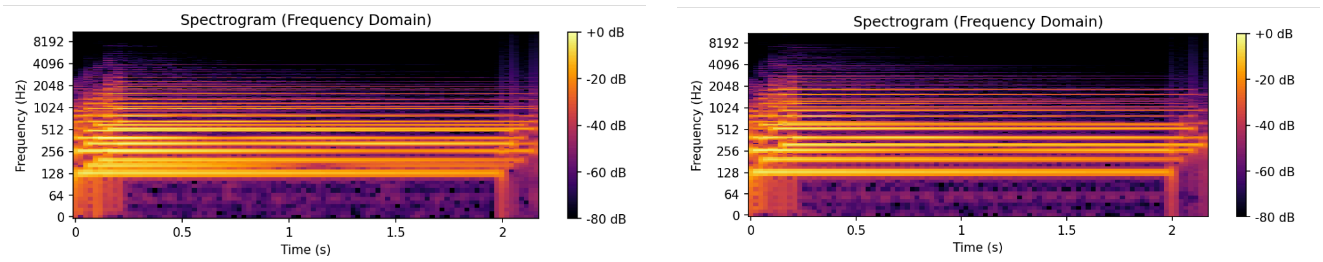


Image 6 and 7: spectrogram of a major and minor signal respectively

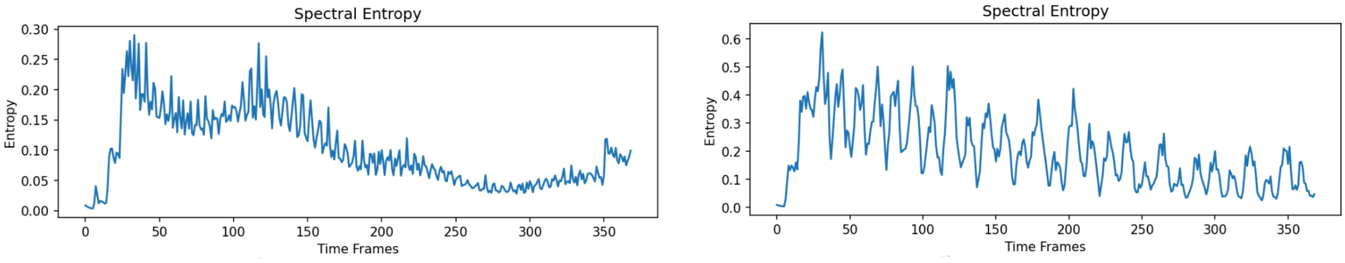


Image 8 and 9: entropy vs time frames of a major and minor signal respectively

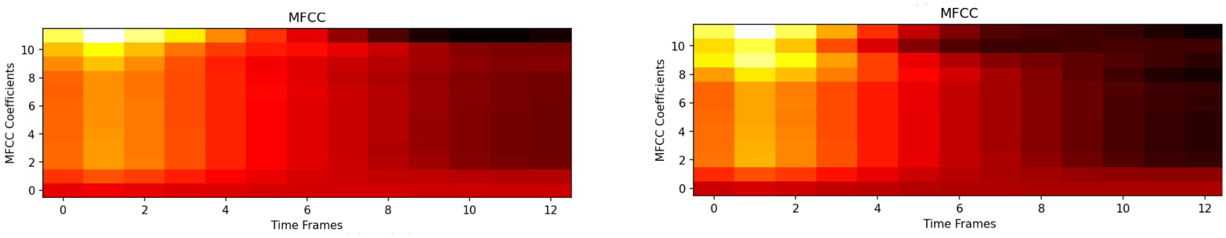


Image 10 and 11: Mel-frequency Cepstral Coefficients for 13 time frames for a major and minor signal respectively

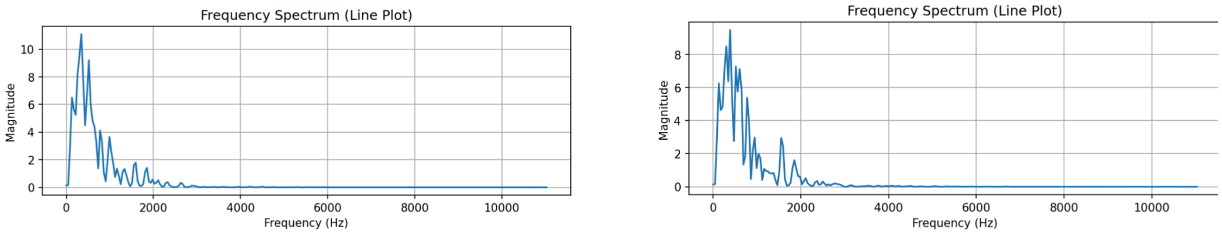


Image 12 and 13: Magnitude Frequency plot for a major and minor signal respectively.

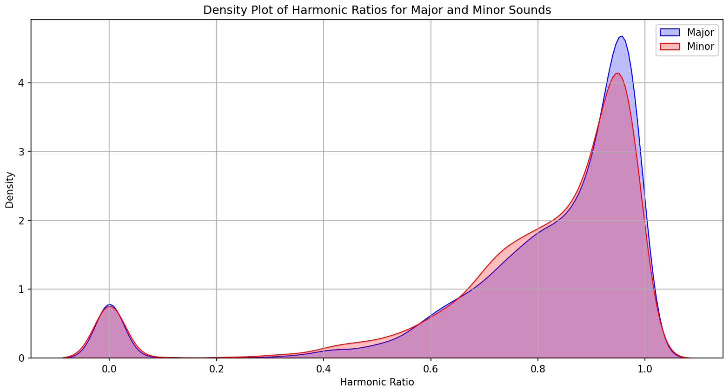


Image 14: density plot of harmonic ratios for major and minor chords

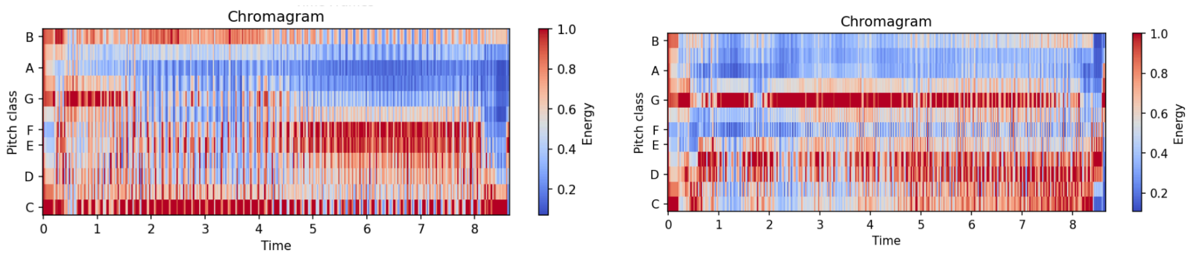


Image 15 and 16: chronogram for major and minor chords respectively

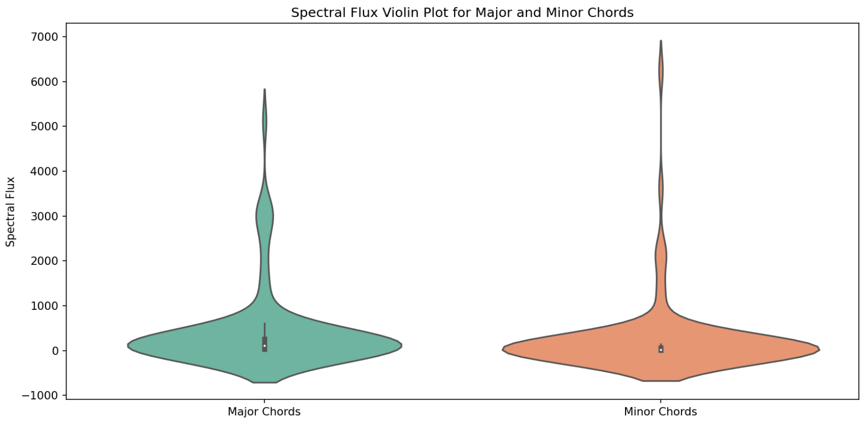
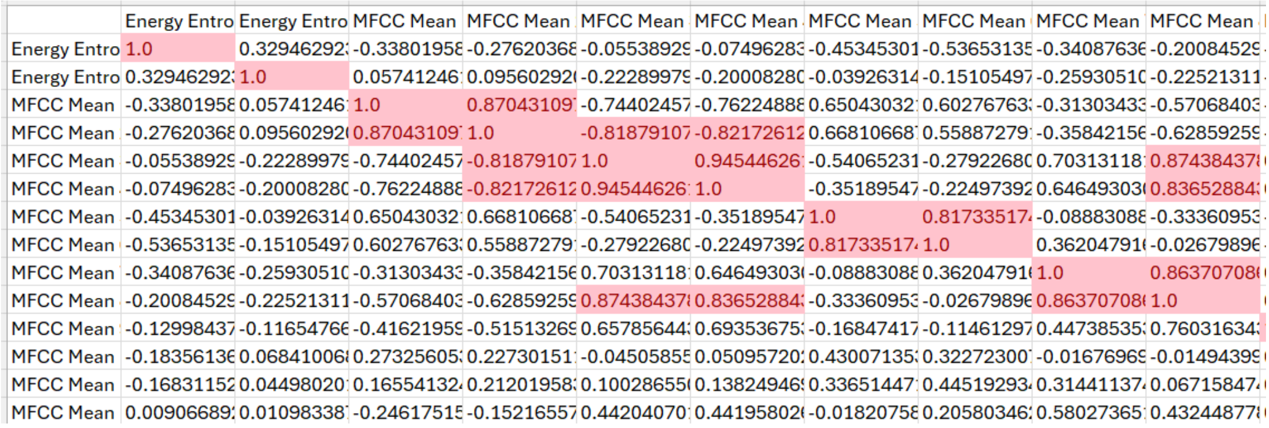


Image 17: violin plot of spectral flux for major and minor chords respectively



	Energy Entro	Energy Entro	MFCC Mean	MFCC Mean	MFCC Mean	MFCC Mean	MFCC Mean	MFCC Mean	MFCC Mean	MFCC Mean
Energy Entro	1.0	0.32946292	-0.33801958	-0.27620368	-0.05538929	-0.07496283	-0.45345301	-0.53653135	-0.34087636	-0.20084529
Energy Entro	0.32946292	1.0	0.05741246	0.09560292	-0.22289979	-0.20008280	-0.03926314	-0.15105497	-0.25930510	-0.22521311
MFCC Mean	-0.33801958	0.05741246	1.0	0.87043109	-0.74402457	-0.76224888	0.65043032	0.60276763	-0.31303433	-0.57068403
MFCC Mean	-0.27620368	0.09560292	0.87043109	1.0	-0.81879107	-0.82172612	0.66810668	0.55887279	-0.35842156	-0.62859259
MFCC Mean	-0.05538929	-0.22289979	-0.74402457	-0.81879107	1.0	0.94544626	-0.54065231	-0.27922680	0.70313118	0.87438437
MFCC Mean	-0.07496283	-0.20008280	-0.76224888	-0.82172612	0.94544626	1.0	-0.35189547	-0.22497392	0.64649303	0.83652884
MFCC Mean	-0.45345301	-0.03926314	0.65043032	0.66810668	-0.54065231	-0.35189547	1.0	0.81733517	-0.08883088	-0.33360953
MFCC Mean	-0.53653135	-0.15105497	0.60276763	0.55887279	-0.27922680	-0.22497392	0.81733517	1.0	0.36204791	-0.02679896
MFCC Mean	-0.34087636	-0.25930510	-0.31303433	-0.35842156	0.70313118	0.64649303	-0.08883088	0.36204791	1.0	0.86370708
MFCC Mean	-0.20084529	-0.22521311	-0.57068403	-0.62859259	0.87438437	0.83652884	-0.33360953	-0.02679896	0.86370708	1.0
MFCC Mean	-0.12998437	-0.11654766	-0.41621959	-0.51513269	0.65785644	0.69353675	-0.16847417	-0.11461297	0.44738535	0.76031634
MFCC Mean	-0.18356136	0.06841006	0.27325605	0.22730151	-0.04505855	0.05095720	0.43007135	0.32272300	-0.01676969	-0.01494399
MFCC Mean	-0.16831152	0.04498020	0.16554132	0.21201958	0.10028655	0.13824946	0.33651447	0.44519293	0.31441137	0.06715847
MFCC Mean	0.00906689	0.01098338	-0.24617515	-0.15216557	0.44204070	0.44195802	-0.01820758	0.20580346	0.58027365	0.43244877

Image 18: partial correlation matrix of all the features for feature selection