EMOTION MUSIC CLASSIFIER

Mariona Carós Roca

Abstract

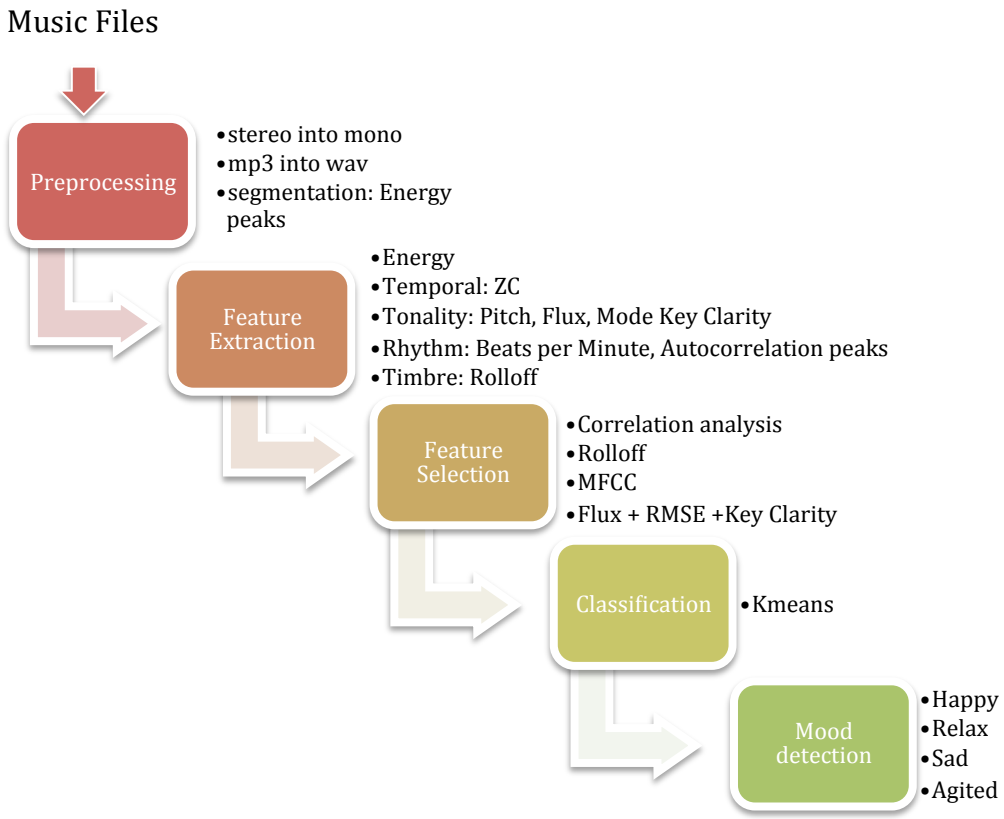
In this report, It is presented an analysis of audio features extracted from music files to associate a song with an emotion. This work is based on 50 songs including movie's soundtracks, popular music, hard rock, etc. Each song is classsified into one of the moods: happy, sad, agited and relaxed with the proposal of creating a playlists for each emotion. Using Music Information Retrieval (MIR) techniques, audio features are extracted to obtain information of tonality, energy, rhythm, and timbre. Having selected the best features, a Kmeans classifier is implemented to enable the classification.

1. Introduction

Technology enables to save enormous volumes of multimedia content to be distributed across the Internet. In order to exploit this data, content providers require effective tools for efficient selection and classification of relevant content. A core technology for such information management tools is the automated analysis of the content to index significant features. Musical aspects definitely play an important role in deciding the emotion of a song. In this study it is proven that emotion can be defined in terms of audio features.

First of all a preprocessing of the audio signal is needed for the feature extraction, which is explained in section 2. On the following sections (3 and 4) it is explained which features were extracted and why and which ones were selected to use in the classifier. In Section 5 it is explained the implementation of K-means algorithm to classify the songs. In section 6 is shown the user interface and the obtained results.

The following digram summarizes the project.



2. Preprocessing

To extract the features I used the *MIRtoolbox* version 1.3.4 which is a MATLAB toolbox that offers an integrated set of functions to extract musical features from audio files.

Firstly, the .mp3 stereo audio files are transformed into mono and saved as a .wav files with a sample frequency of 44.100 Hz to be compatible to the toolbox. Each song could contain a range of different emotions, so a segment of 40 seconds is extracted from the audio signal. Specifically between second 10 and 60, becuase in some track files the first seconds are in silence. I though that 40 seconds was informative enough to retrieve the mood of the whole song.

Then the audio signal is segmented using the function *mironsets* which detects the peaks of the curve indicating the point where the energy is the highest, so we can suppose that these peaks corresponds the successive notes in the music. This way segments are more stationary and easy to extract valuable information. The diference between classic segmentation and *mironsets* segmentation is shown in the following images.

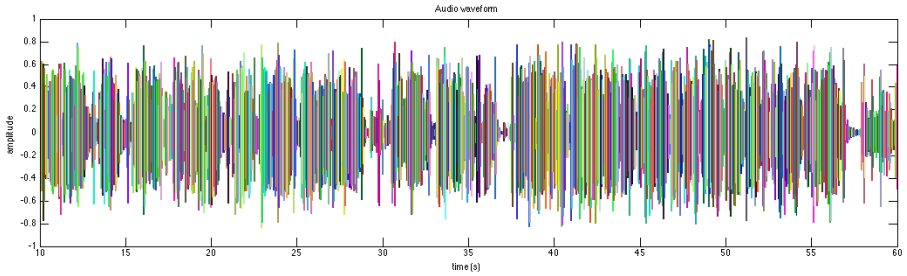


Figure 1. Segmentation with same length frame

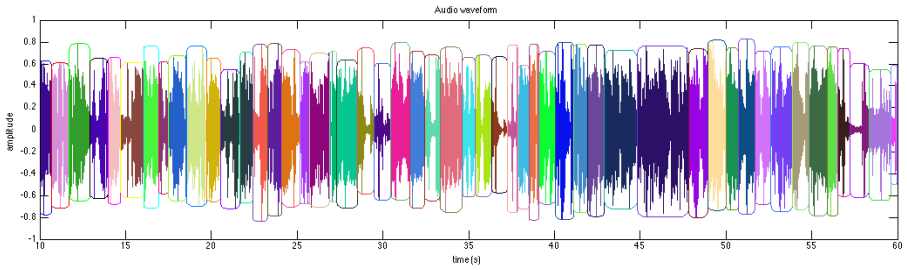
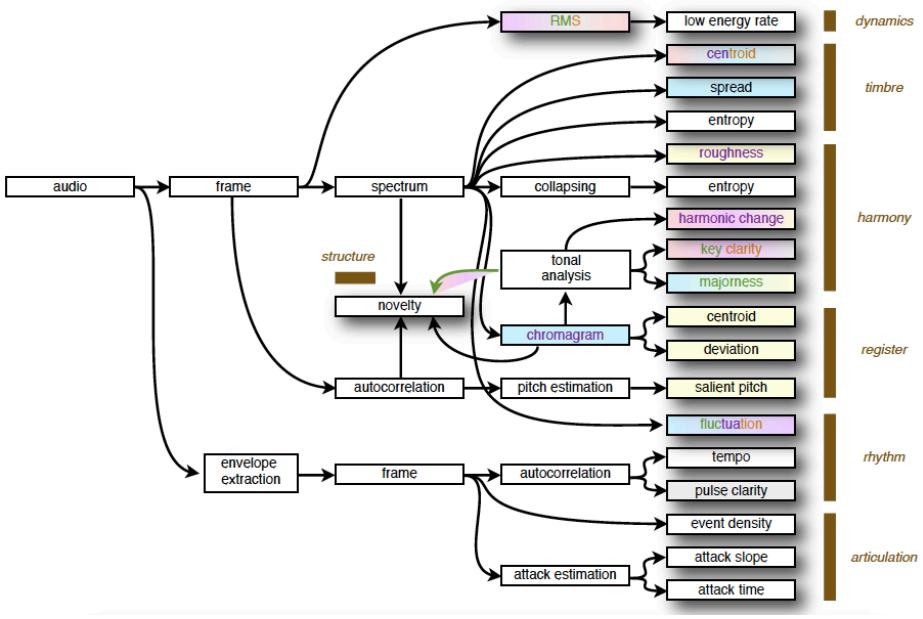


Figure 2. Segmentation analysing the peack atack energy.

3. Feature extraction

Audio features about the mode, ryhthm, tonality, timbre and energy of music were extracted using MIRToolbox. The diagram shows the features extracted from audio files using MIRToolbox.



3.1. Features:

Energy: Indicator of activity, as much energy more agited will be the song. The short-time energy of an audio signal is defined as:

$$\mathrm{En} = \frac{1}{\mathrm{N}} \sum_{\mathrm{m}} [\mathrm{x(m)w(n - m)}]^2$$

Where x(m) is the sampled audio signal, n is the time index of the short-time energy, and w(m) a rectangular window. En is calculated for each frame.

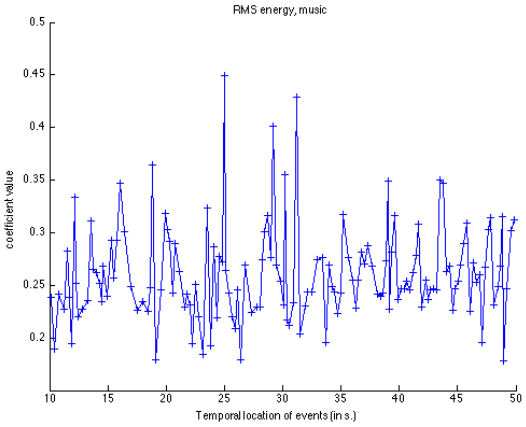


Figure 3. RMS Energy of happy song

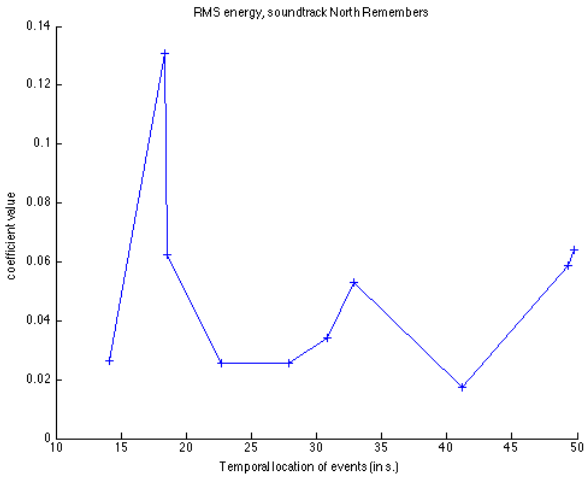


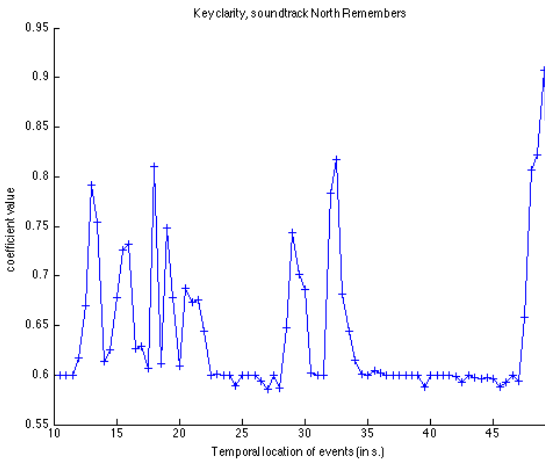
Figure 4. RMS Energy of sad song

Rhythm: is the pattern of strong and weak beat. It can be described through speed (tempo), strength, and regularity of the beat

- Ryhtm regularity: autocorrelation of spectrum
- Zero crossing rate: number of times the signal crosses zero line
- BPM (Beats Per Minute).

Tonality: These features included the most probable key of the music as well as an estimation of whether the key was major or minor.

- Key Clarity and Roughness
- Mode: It estimates if the piece is in major, or in minor mode



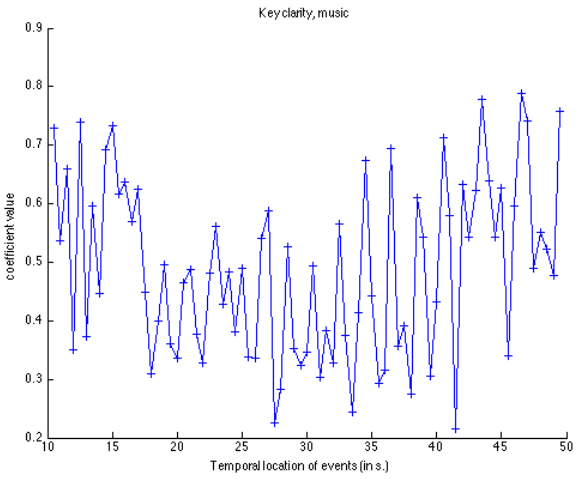


Figure 5. Key Clarity of sad song (soundtrack) *Figure 6. . Key Clarity of angry song*

We can see at the figures that keys are much more clear at a sad song that in a hard rock song for example.

Articulation: It refers to the transition and continuity of the music, it is an indicator of how aggressively a note is played.

- Attack slope: Peaks of spectral changes

Timbre: describes the quality of the sound. It is often defined in terms of features of the spectrum gathered from the audio signal. All these features are taken from the signal spectrum.

- Brightness: Amount of energy above a cutoff point in the spectrum
- Rolloff: The frequency such that 85% of total energy is contained below that frequency
- Spectral flux: Average distance between the spectrum of successive frames spectral centroid (the frequency around which the spectrum is centered
- MFCC: Rrepresentation of the short-term power spectrum of a sound separated into different bands. A way to represent time domain waveforms with just a few frequency domain coefficients
- Spectral Centroid: It is defined as the center of gravity of the spectrum
- Flux: Spectral changes in time

On the following images we can see that Rolloff is a feature that varies a lot from one emotion to another. It has many points to be one of the selected.

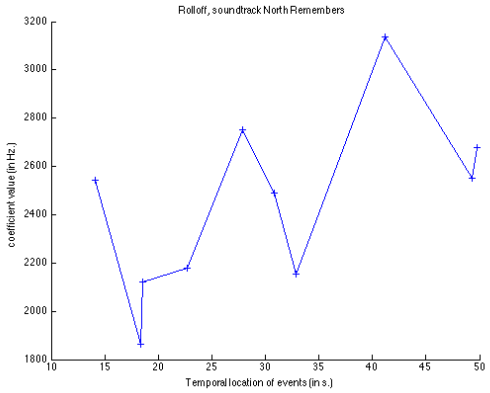


Figure 7. Rolloff of sad song

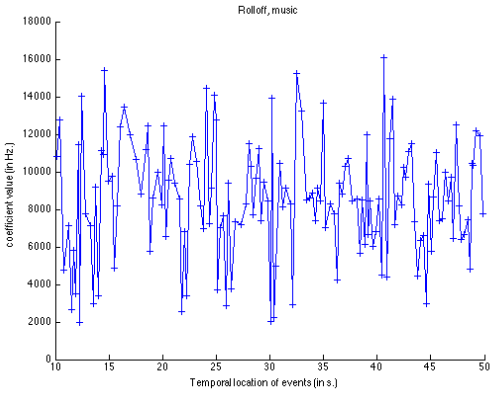


Figure 8. . Rolloff of happy song

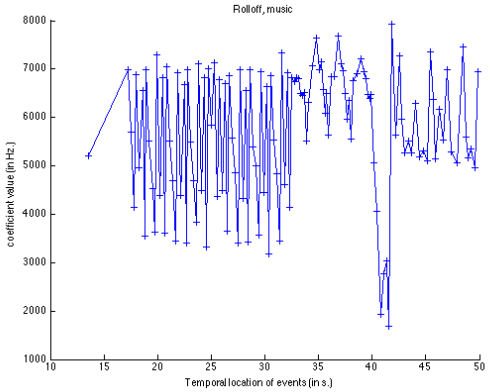


Figure 8. Rolloff of angry song

4. Feature Selection

It is important to select a particular set of features to get best classification results and improve computational cost. The selected set would determine the nature of the class, so it has to be conscientiously chosen. The number of features determine the dimensionality in the feature space, I needed to chose as maximum 3, because I wanted to plot the results.

Therefore to make the decision I took into account:

Uncorrelated to other features: It is very important that there are no redundancies in the feature space. Each new feature that is selected must give altogether different information about the signal as possible. This helps in better computation efficiency, improved performance and optimization of cost.

So I computed the correlation matrix of all features and I saw that there were so many correlated between them.

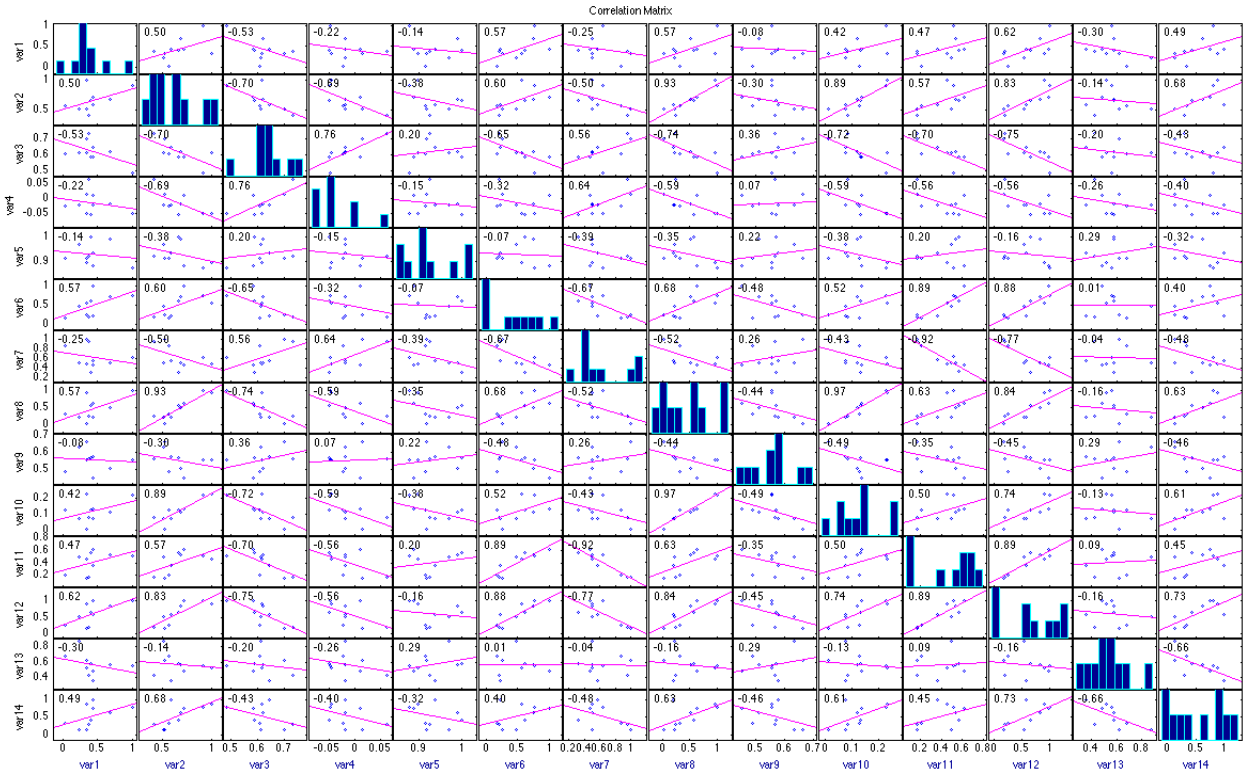


Figure 9 . Correlation matric of 14 features

On the following figure we can se a clearly correlation between the ploted features ZC, fluctuation, brightness and flux.

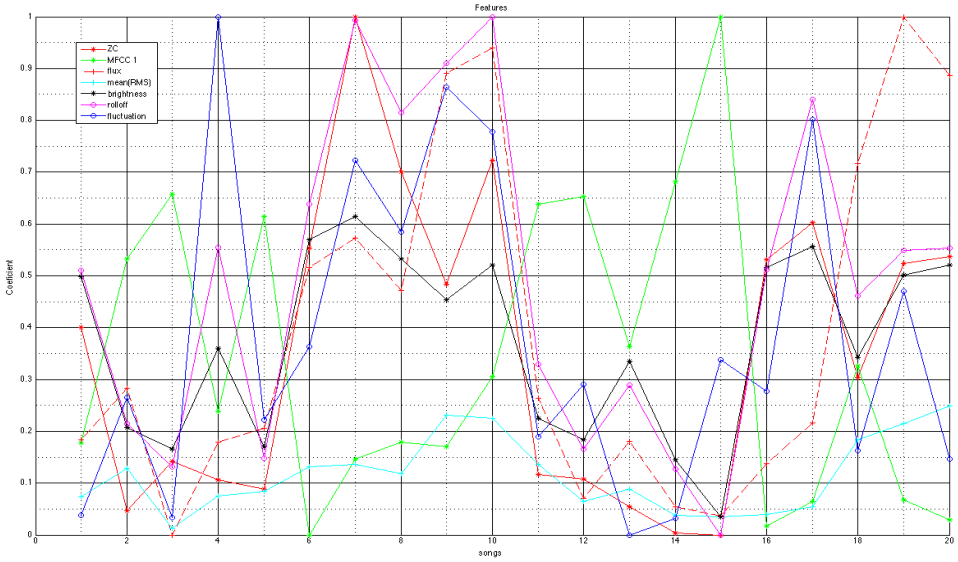


Figure 9 . Correlation matric of 14 features

Discriminative Power: The purpose of feature selection is to achieve discrimination among different classes of audio patterns. Therefore a feature must take round about similar values within the same class but different values across different classes, as we can see on the previous image, where songs between 1 - 5 are sad, 6 - 10 are happy, 11 - 15 are relaxes and 16 - 20 are agited and the feature values are clearly different.

Invariance to irrelevancies: Any good feature should exhibit invariance to irrelevancies such as noise, bandwidth or the amplitude scaling of the signal.

Finally, I selected the features: Rolloff, MFCC(1), RMS energy, flux, Key clarity and Rhythm regularity beacuse they weren't correlated between them and they were the less invariance in their emotion class and the most discriminant between other classes.

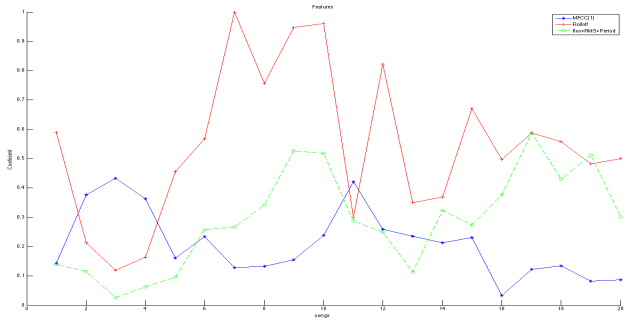


Figure 9. Chosen features

5. Classification

Classification is the process by which a particular label is assigned to a particular cluster. It is this label that would define its class, its emotion. I chose K-means beacuse it is simple and it obtains good results.

The goal of the algorithm is to cluster the data base into 4 classes, each one corresponding to one emotion. Centroids are initializated with determined values which I computed with the train data to obtain the feature's mean of all vectors from each class. The criterion of classification is the Euclidean distance, so each vector of the data base will pertain to the cluster of the nearest centroid. For every vector the euclidian distance is calculated with all centroids and then the vector is labeled. When the distance has been calculated for all database centroids are recalculated and the algorithm iterates successively until the centroid no longer change its value.

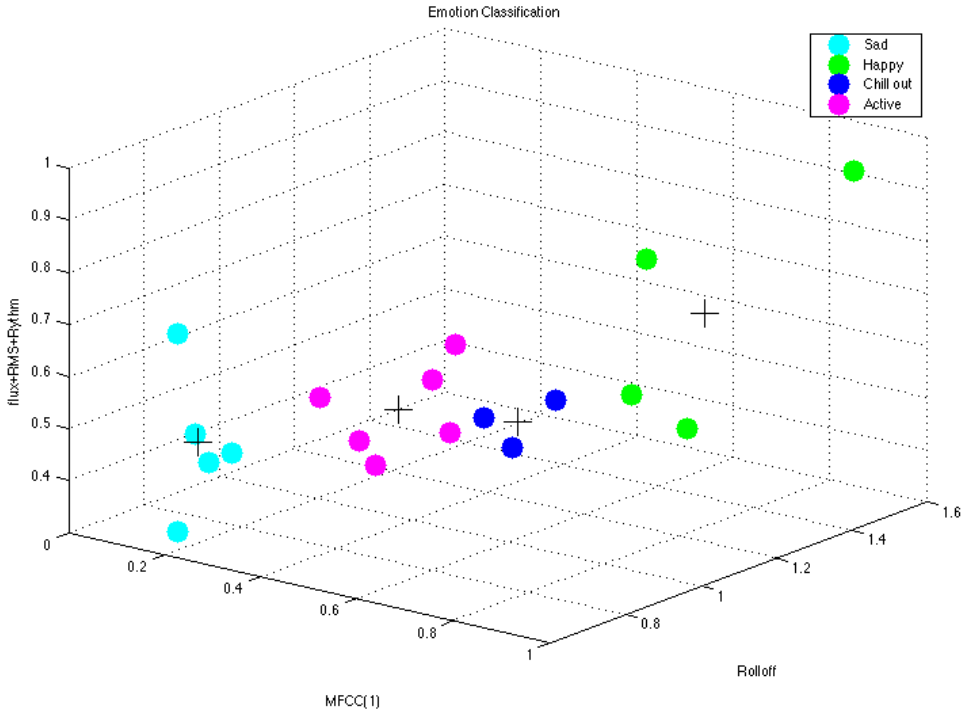


Figure 10. Mood Classification Plot by K-means algorithm

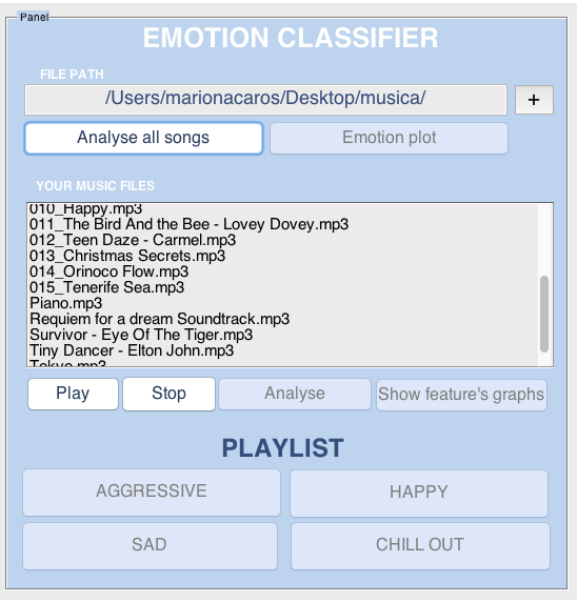
6. Results

6.1 Classification

Music File	k-means	Class
'001_Titanic.mp3'	4	1
'002_MadWorld.mp3'	1	1
'003_Spiegel.mp3'	1	1
'004_The North Remembers.mp3'	1	1
'005_Now We Are Free.mp3'	1	1
'006_Get Lucky.mp3'	2	2
'007_Drop In The Ocean.mp3'	2	2
'008_Gang Of Rhythm.mp3'	2	2
'009_Lets Go Surfing.mp3'	2	2
'010_Happy.mp3'	2	2
'011_The Bird And the Bee - Lovey Dovey.mp3'	1	3
'012_Teen Daze - Carmel.mp3'	3	3
'013_Christmas Secrets.mp3'	3	3
'014_Orinoco Flow.mp3'	3	3
'015_Tenerife Sea.mp3'	3	3
'016_Killing In The Name.mp3'	4	4
'017_Eminem - Not Afraid.mp3'	4	4
'018_Zeds Dead ft Omar Linx-Out For Blood.mp3'	4	4
'019_Battle Cry.mp3'	4	4
'020_Fine Without You with Lyrics.mp3'	4	4

6. 2 User interface

The user interface is intuitive and easy to use. When the application starts the user must input the path of the folder cointaining the music and click de button "+". Then he can play a song of his library or analyse the whole folder. When the analysis finishes more buttons are enabled. Then the user can see the plot his songs classification or obtain his mood playlist. He can also see the features such as Energy, Rolloff, spectrum... of a specific music file by clicking "Show feature's graph".



7. Conclusions

The main important variable of the project is "features". It is important to select an optimum number of features that not only keeps accordance with the accuracy and the level of performance but also reduces the computation costs.

Songs are classified into emotion categories based on weighted combination of features chosen accurately. This combination is achieved by experimentation. Finally the most important features for emotion classification are in the first place MFCC for its spectral information compressed in few coefficients, and Rolloff for its discrimination among different classes, in second place: flux, RMSE and Key clarity.

The main proposal of this classification algorithm is to generate personal playlists using the user's music library. This way the user would have a playlist for training or running with his songs, another for moments of concentration such as study or read also made with his music.

8. References

MIRToolbox Manual 1.3.4
MIRToolbox Premier
L. Lu, D. Liu, and H.J. Zhang, "Automatic mood detection and tracking of music audio signals," Audio, Speech, and Language Processing, IEEE Transactions Jan. 2006.
G. Tzanetakis and P. Cook, Musical Genre Classification of Audio Signals, IEEE Trans. Speech and AudioProcess, vol. 10, pp. 293 302, 2002 July.

9. Testing

Computer specifications:
MATLAB Version: 8.3.0.532 (R2014a)
MATLAB License Number: 405329
Operating System: Mac OS X Version: 10.11.6 Build: 15G1108
Java Version: Java 1.7.0_11-b21 with Oracle Corporation Java HotSpot(TM) 64-Bit Server VM mixed mode

Used toolbox:
MIRtoolbox
AuditoryToolbox

Downloaded at:
https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox