UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR

TRABAJO FIN DE MÁSTER

Characterisation and analysis of emotions from musical stimuli in biological signals

Máster Universitario en Ingeniería Informática e I2-TIC

Autor: TIRADO LÓPEZ, Laura Olga

Tutor: VARONA MARTÍNEZ, Pablo

Departamento de Informática

Junio de 2019

Characterisation and analysis of emotions from musical stimuli in biological signal

Laura Olga Tirado López

Keywords: computational neuroscience, emotion recognition, feature extraction, emotion classification, EEG, affective computing

Abstract

Emotions and the brain activity behind them is a subject extensively addressed in recent years, whereas is related to more physiological fields or therapeutic ones. Emotions can be elicited by many things: pictures, memories, words or sounds, to name a few. The latter, in the form of music, are one of the most engaging ones.

Music as stimuli for emotion recognition is challenging and provocative, not only because of its complexity as a signal, but for its many applications related to mental therapies and its relation to memory brain process.

In the present study, we analyse and characterise different biological signals, mostly EEG signals, in an attempt to classify emotions triggered by music stimuli. We implement different feature selection methods, based on grid search with cross-validation, and machine learning algorithms, both supervised and unsupervised learning, to address the effect of musical emotion. Moreover, we try different sets of characteristics and sampling rates to validate how explanatory are the selected features.

Additionally, we analyse the stimuli for the purpose of unveiling how musical features are related to certain emotions in terms of valence and arousal.

Caracterización y análisis de emociones a través de estímulo musical en señales biológicas

Laura Olga Tirado López

Palabras clave: neurociencia computacional, reconocimiento de emociones, extracción de características, clasificación de emociones, EEG, computación afectiva

Resumen

Las emociones y la actividad cerebral detrás de éstas es un tema que ha sido ampliamente estudiado en los últimos años, ya sea relacionado con campos de estudios fisiólogico o terapeúticos. Las emociones pueden ser suscitadas por muchas cosas: imágenes, recuerdos, palabras o sonidos por mencionar algunos. Éstos últimos, en forma de música, son uno de los más interesantes.

La música como estímulo para el reconocimiento de emociones es exigente y desafiante, no sólo por su complejidad como señal, sino por sus múltiples aplicaciones relacionadas con salud mental y su relación con procesos cerebrales de la memoria.

En este proyecto, analizamos y caracterizamos distintas señales biológicas, principalmente señales EEG, en un intento de clasificar emotiones provocadas por estímulos musicales. Se implementan diferentes métodos de selección de características, basados en búsqueda grid con validación cruzada, y algorimos de aprendizaje autómatico, tanto de aprendizaje supervisado como no supervisado, para estudiar el efecto de las emociones provocadas por la música. Además, se emplean diferentes conjuntos de características y frecuencias de muestreo para validar cómo de explicativas son las características seleccionadas.

Adicionalmente, se analiza el estímulo con el propósito de desvelar cómo las características musicales se relacionan con ciertas emociones en términos de valencia y estimulación.

Acknowledgements

To my family for their endless patience, immense support, unconditional love and make it possible for me to be who I am today.

To my friends, old and new ones, for every smile, laugh and moment we share.

To my dearest friend Isa. Thank you for being with me in the distance and being my Jiminy Cricket.

To my thesis director, Pablo, for bring me the opportunity to make this project, for his support, motivation and immense knowledge.

And lastly but not least, to my girls, for getting the best of me.

Contents

List of Figures

List of Tables

1. Introduction and State of the Art

This chapter present the motivation, the objectives and the structure of this thesis. It is further structured as follows: Section 1.1 presents the motivation of this work, Section 1.2 states the objectives of this thesis and Section 1.3 outlines the structure of this thesis by presenting an overview of the chapters.

1.1. Motivation

The study of emotions and their relationship with brain activity is a subject widely studied with many possible applications: from a better understanding of the brain and its functioning, to the creation of various therapies and treatments for mental diseases. For the latter, emotion classification or emotion recognition is an emerging topic, especially in the biomedical context.

There are several studies which present various classification models for emotion recognition, using biological signals data as input, usually EEG signals, e.g. see [Jenke et al., 2014, Jirayucharoensak et al., 2014, Zheng and Lu, 2015, Jatupaiboon et al., 2013]. EEG signals allow us to see the brain activity as time series and identify certain events.

However, the analysis of multivariate biological signals is very challenging due to its noisy nature, temporal heterogeneity, drift, etc. which require taking into account the specific characteristics of these time series [Pourahmadi and Noorbaloochi, 2016, Rozado et al., 2010, Rozado et al., 2012b, Rozado et al., 2012a, Baydogan and Runger, 2015]. In particular, EEG signals have a lot of noise and artefacts, that are hard to remove without loosing information. Moreover, these signals vary from one individual to another, which means there is not a proper “general shape” to describe them.

One of the most interesting studied topics related to emotion recognition is with musical stimuli, which as it is exposed in [Jäncke, 2008] seems to be also related to memory functions.

Music has a prominent role in the everyday life of many people. Whether it is for recreation, distraction or mood enhancement, people are constantly exposed to this stimuli. Music has several similarities to speech, mainly structural ones since we can define music as a language on its own. However, music is a far more complex signal to process than speech. In the context of emotion recognition from biological signals, music is an adequate stimulus due to its temporal structure and the emotional charge that we can attribute to it.

One of the main goals in this project is to characterise music evoked EEG signals, along others biological signals (temperature and breathing rhythm), and find any relevant feature which lead us to identify certain emotions.

Besides that, there is not too much research about how musical stimuli relates to emotions from neural recordings, although some experimental and theoretical studies have addressed this topic (e.g., see [Purwins et al., 2008b, Purwins et al., 2008a, Song et al., 2012, Salimpoor et al., 2015, Varona and Rabinovich, 2016]). This could be helpful in terms of identifying certain music features that lead us to feel certain emotions, which can be applied in certain biomedical fields, like music therapy.

1.2. Objectives

Based on the motivation given above, this project have two main objectives:

- **Analysis and classification of emotions using biological signals**: characterise and extract features from this signals in order to implement a model for emotion recognition or emotion classification. The signals that we will be using for this purpose will be EEG recordings, temperature an breathing rhythm.

- **Analysis and feature extraction of the stimuli**: analyse the stimuli in order to find any relevant feature or features that can be related to the feeling of any specific emotion. For this purpose, we will analyse the stimuli in terms of music features like tone, mode or tempo among others.

1.3. Report structure

This report contains four chapters detailed below:

- Chapter 1 (Introduction) presents the motivation, objectives and structure of this project.

- Chapter 2 (Design and Development) presents the data acquisition process and the methodology applied in both emotion classification and stimuli analysis.

- Chapter 3 (Results) presents the results obtained with the feature extraction and classification models presented in chapter 2.

- Chapter 4 (Conclusions) presents the discussion of the results shown in chapter 3, final conclusions and future work.

2. Design and Development

2.1. Data acquisition and dataset

Taking into account the goals described in the previous chapter, we searched several EEG databases focused on emotion classification finding the DEAP dataset [Koelstra et al., 2012] the most relevant one for our analysis purposes.

This multivariate dataset contains EEG data and other biological data like eye-tracking, temperature and breathing rhythm under musical and visual stimulation. To access to this dataset we contacted the administrator.

DEAP is a multimodal dataset for the analysis of human affective states. It contains EEG and peripheral physiological signals (eye-tracking, temperature, breathing rhythm, ...) of 32 participants recorded while watching music videos. Participants watched a total of 40 music videos for one minute each and rated them in terms of levels of arousal, valence, liking, dominance and familiarity.

Trials were realised with 40 videos playing in a different order for each participant. However, the authors kept the data organised in the dataset to avoid mistakes while using it for research purposes.

The emotional rating of the videos was made using self- assessment manikins (SAM) graphics [Bradley and Lang, 1994], which are showed in figure 2.1.1. They were rated by selecting the most suitable graphic for each trial in a continuous scale from 1 to 9.

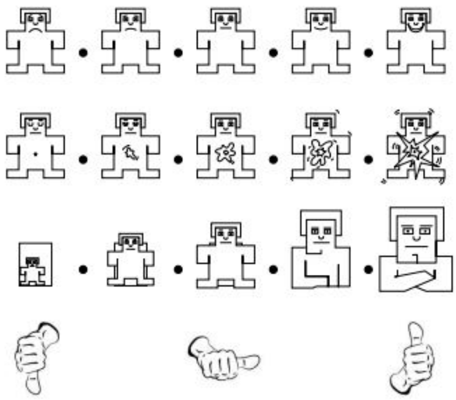

Figure 2.1.1: SAM graphics used for selfassessment rating. Image from [Koelstra et al., 2012].

The figure below shows the files of the dataset:

File name	Format	Part	Contents
Online_ratings	xls, csv, ods spreadsheet	Online self-assessment	All individual ratings from the online self-assessment.
Video_list	xls, csv, ods spreadsheet	Both parts	Names/YouTube links af the music videos used in the online self-assessment and the experiment + stats of the individual ratings from the online self-assessment.
Participant_ratings	xls, csv, ods spreadsheet	Experiment	All ratings participants gave to the videos during the experiment.
Participant_questionnaire	xls, csv, ods spreadsheet	Experiment	The answers participants gave to the questionnaire before the experiment.
Face_video	Zip file	Experiment	The frontal face video recordings from the experiment for participants 1-22.
Data_original	Zip file	Experiment	The original unprocessed physiological data recordings from the experiment in BioSemi .bdf format
Data_preprocessed	Zip file for Python and Matlab	Experiment	The preprocessed (downsampling, EOG removal, filtering, segmenting etc.) physiological data recordings from the experiment in Matlab and Python(numpy) formats

Figure 2.1.2: Dataset files.

For our project we have used the data from the following files:

- *Video_list.xls*: this file lists all the videos used in the experiment and their related information: experiment id, artist, title and average values of valence, arousal and dominance.

- *Data preprocessed*: these files contain all EEG and physiological data recordings from the experiment preprocessed. To obtain these preprocessed data, authors downsampled the original data to 128Hz, removed EOG artefacts, applied a band-pass frequency filter from 4.0Hz to 45.0Hz, averaged the data to the common reference and segmented it into 60 second trials.

Each file contains 32 `.dat` files, one per participant, with two arrays:

- data: which contains a 40x40x8064 (trial x channel x data) array with the data of each trial and EEG channel.
- labels: which contains a 40x4 (trial x label) array with the values of valence, arousal, dominance and liking rated by the participant for each video.

The channels layout are organised as described below:

- Channels 1 to 32: EEG channels.
- Channels 33 to 37: EOG channels, which track the eye-movements by means of electrooculography.
- Channels 38 to 40: respiration belt, plethysmograph and temperature.

For this project, we used the information provided by EEG, respiration and temperature channels.

2.1.1. EEG

Electroencephalography or EEG is an electrophysiological monitoring method to record electrical activity of the brain. This method is typically noninvasive, with electrodes placed along the scalp. Each electrode is connected to one input of an amplifier, which records the activity of huge populations of brain cells.

Electrodes locations are specified by the International 10-20 system 2.1.3. This system ensures that the naming of electrodes is consistent across laboratories. In most clinical applications, 19 recording electrodes (plus ground and system reference) are used. However, additional electrodes can be added to the standard set-up.

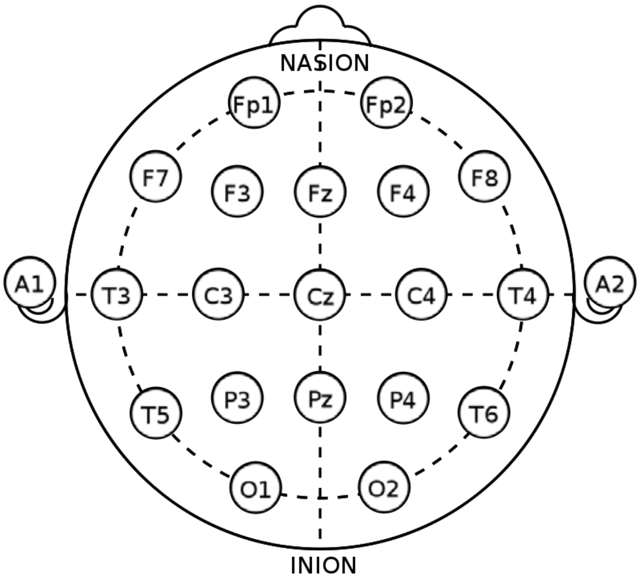

Figure 2.1.3: International 10-20 system.

EEG refers to the recording of the brain's spontaneous electrical activity over a period of time, which means that EEG data is composed of multiple time series, one per electrode. For the analysis, EEG rhythmic activity is typically divided into bands by frequency. These frequency bands are a matter of nomenclature which have a certain distribution over the scalp or a certain biological significance [Jenke et al., 2014]. The typical frequency bands used in EEG studies are listed below:

- **Delta**: its frequency range is up to 4Hz. It is seen normally in adults in slow-wave sleep and babies.

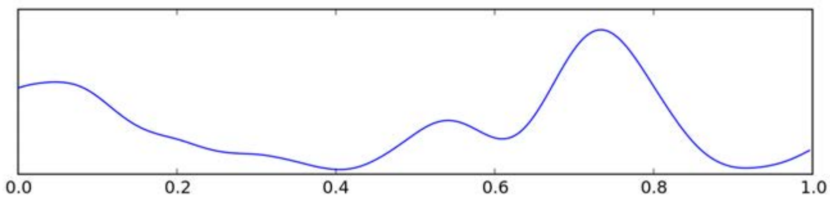

Figure 2.1.4: Delta wave.

- **Theta**: its frequency range is from 4Hz to 7Hz. It is seen normally in young children and may be seen in drowsiness or arousal in adults and also in meditation. This range is associated with relaxes, meditative and creative states.

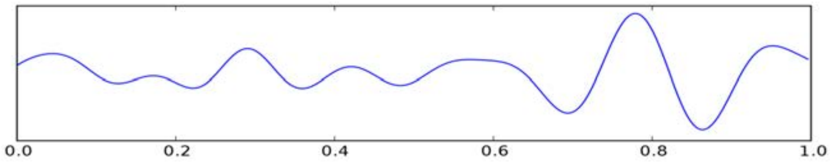

Figure 2.1.5: Theta wave.

- **Alpha**: its frequency range is from 7Hz to 13Hz. It is seen in the posterior regions of the head on both sides. It emerges with closing of the eyes and with relaxation, and attenuates with eye opening or mental exertion.

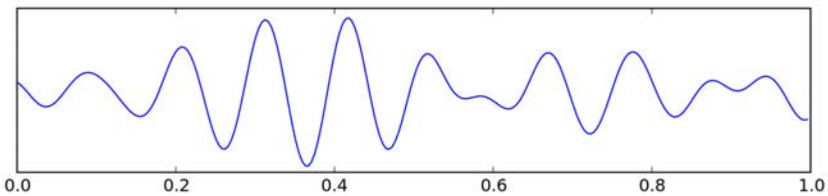

Figure 2.1.6: Alpha wave.

- **Beta**: its frequency range is from 14Hz to 30Hz. It is seen usually on both sides in symmetrical distributions and is mos evident frontally. It is closely linked to motor behaviour.

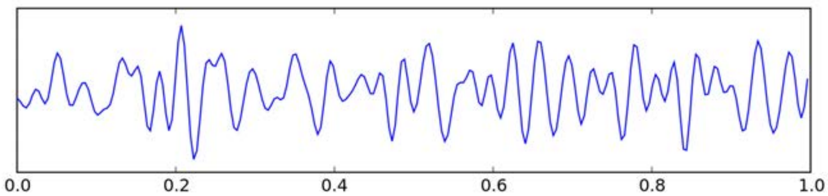

Figure 2.1.7: Beta wave.

- **Gamma**: its frequency range is from 30Hz to 100Hz approximately. Gamma rhythms are though to represent binding of different populations of neurons together into a network for the purpose of carrying out a certain cognitive or motor function.

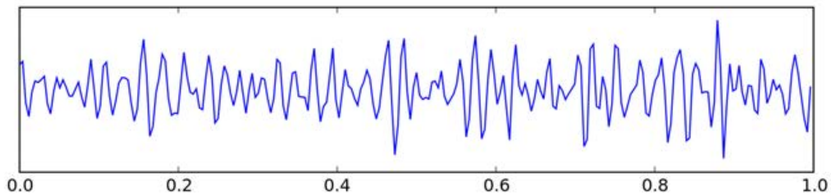

Figure 2.1.8: Gamma wave.

2.1.2. Circumplex Model of Emotion

We have explained the nature of our data, yet we need to answer one question: how are we going to label it? In other words, how are we going to define emotions. Since we have valence and arousal values in our dataset, we will be using the circumplex model of emotion.

The circumplex model of emotion was developed by James Russell [Posner et al., 2005]. This model suggest that emotions are distributed in a two-dimensional circular space, containing arousal and valence dimensions. Arousal is represented in the vertical axis and valence is represented in the horizontal axis, while the centre of the circle represents a neutral valence and a medium level of arousal. In this model, emotional states can be represented at any level of valence and arousal. According to this model, all affective states are the product of two independent neurophysiological systems: valence-neural circuitry and arousal-neural circuitry.

The valence-neural circuitry relates to the mesolimbic system, which have long been associated with pleasure. Therefore, the valence dimension represents the level of liking or disliking.

The arousal-neural circuitry relates to the limbic system, the thalamus and the amygdala, where it have been suggested that the neural representations of the emotional significance reside. The arousal dimension represents how a stimuli affects these systems in terms of activation: active or passive responses.

As it is shown in figure 2.1.9, our representation space is divide in four main sections according to valence-arousal values, which represents four main emotions: excitement, distress, calm and sorrow.

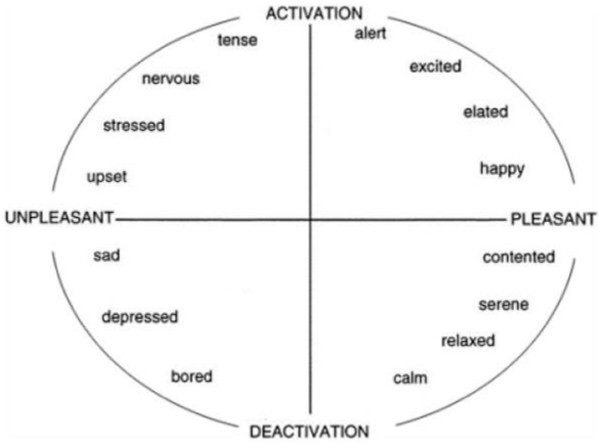

Figure 2.1.9: Graphical representation of the Circumplex model of affect. Image from [Posner et al., 2005].

We can describe these four sections or quadrants as below:

- Right upper quadrant: HVHA section (high valence and arousal values). Excitement related emotions locate in this quadrant like happiness or elation.

- Left upper quadrant: LVHA section (low valence and high arousal values). Distress related emotions locate in this quadrant like stress or nervousness.

- Right lower quadrant: HVLA section (high valence and low arousal values). Calm related emotions locate in this quadrant like serenity or content.

- Left lower quadrant: LVLA section (low valence and low arousal values). Sorrow related emotions locate in this quadrant like sadness or boredom.

In the DEAP dataset, valence and arousal values are represented in the range [1-9].

2.2. Methodology

This section presents in detail the applied methodology in order to achieve the main objectives of this project in two distinct sections:

- **Emotion classification**: Feature extraction and emotion classification of biological signals of the DEAP dataset.

- **Stimuli analysis**: Characterisation and analysis of the stimuli used for emotion recognition.

2.2.1. Emotion classification

As it is explained in the preceding section, the DEAP dataset provides various biological signals: EEG, EOG, breathing rhythm, plethysmograph and temperature. We decided to use EEG, breathing rhythm and temperature out of all the signals in the dataset to perform our feature extraction and classification models. Since we wanted to focus on the listening part of the stimuli we discarded the EOG data, which is more related to the visual aspects of the music videos.

2.2.1.1 Feature extraction

Firstly, a good significant representation of the data was needed in order to improve the performance of the classification models. For that purpose, we extracted various features from the data and created feature vectors.

This feature extraction was performanced in various domains since we were working with time series data.

Time domain

As it refers, time domain features are related to the analysis of the signal data with respect to time. The time domain features we used are:

- **Signal mean value**: the average value of the signal.

- **Signal kurtosis value**: a descriptor of the shape of a probability distribution. It is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution. Datasets with high kurtosis tend to have heavy tails, or outliers. Datasets with low kurtosis tend to have light tails, or lack of outliers. We can express it as

$$k = \frac{\sum(X-\mu)^4}{n\sigma^4} - 3 \tag{2.2.1}$$

where X denotes the sequence of inputs, μ represents the mean value of X, σ is referred to the variance of X and n the length of input sequence X.

- **Hjorth features**: indicators of statistical properties used in signal processing in the time domain. These parameters are normalised slope descriptors (NSDs) used in EEG.
  - *Activity*: represents the signal power, the variance of a time function. This can indicate the surface of power spectrum in the frequency domain.

$$Activity = var(y(t)) \tag{2.2.2}$$

  - *Mobility*: represents the mean frequency of the proportion of standard deviation of the power spectrum.

$$Mobility = \sqrt{\frac{var(\frac{dy(t)}{dt})}{var(y(t))}} \tag{2.2.3}$$

  - *Complexity*: represent the change in frequency. The parameter compares the signal's similarity to a pure sine wave, where the value converges to 1 if the signal is more similar.

$$Complexity = \frac{Mobility(\frac{dy(t)}{dt})}{Mobility(y(t))} \tag{2.2.4}$$

Frequency domain

In section 2.1.1, it is explained how EEG signals can be divided into bands by frequency. By its definition, the most interesting bands for our study are the beta and gamma frequency bands. These band are related to wakings consciousness and complex cognitive processes, which appear when an individual is awake and exposed to several stimuli. This is the case of the participants of the DEAP dataset.

We compute the average power of the signal in those specific frequency range (12.5 to 30 Hz for β band and 30 to 100Hz for γ band) using Welch's method.

If a signal $x(t)$ has a Fourier transform $X(f)$, its power spectral density is

$$|X(f)|^2 = S_X(f) \tag{2.2.5}$$

where we defined $S_X(f)$ as $\hat{S}_x^W(\omega_k)$ by the definition of Welch's method explained in [Smith, 2011]:

$$\hat{S}_x^W(\omega_k) \triangleq \frac{1}{K} \sum_{m=0}^{K-1} P_{x_m, M(\omega_k)} \tag{2.2.6}$$

We calculate both absolute power and relative power as features of the signal defined by:

$$\text{Absolute Spectral Power in Band} = \int_{-max}^{-min} \hat{S}_x^W(\omega_k) d\omega_k + \int_{min}^{max} \hat{S}_x^W(\omega_k) d\omega_k \tag{2.2.7}$$

$$\text{Relative Spectral Power in Band} = \frac{\int_{-max}^{-min} \hat{S}_x^W(\omega_k) d\omega_k + \int_{min}^{max} \hat{S}_x^W(\omega_k) d\omega_k}{\int_{-\inf}^{\inf} \hat{S}_x^W(\omega_k) d\omega_k} \tag{2.2.8}$$

where min and max are the min and max values of the frequency intervals for each band ($[12.5, 30]$ for β band and $[30, 100]$ for γ band).

2.2.1.2. Feature selection

In the previous section, we presented the feature extraction process in different domains to build derived values intended to be informative and non-redundant.

However, since biological signals (especially EEG signals) are extremely complex and noisy, we need feature selection to create a subset of the most relevant features for their use in the model construction. This allows us to simplify models, to shorter training times and to enhance generalisation by reducing overfitting risks. We performed this technique using grid search.

Grid search is the process of performing hyper parameter turning in order to determine the optimal values for a given model. This method provided us the best estimator along with features importance, which allowed us to perform feature selection.

Furthermore, we create several sets with selected points of our EEG signals. Since the data sampling rate is 128Hz, we have data records every 0.008 seconds. This is a very small time interval which may not be informative because musical stimuli do not have relevant emotion changes in such a short period of time. In order to test how the sampling rate affected signal information we created subsets of EEG signals selecting one of every 2 to 64 points, changing sampling rates from 128Hz (0.008 seconds) to 2Hz (0.5 seconds).

Lastly, we also reduced the electrodes number. As it is explained in [Valenzi et al., 2014], electrode set reduction facilitates data evaluation. We used the same pool electrodes they proposed, reducing from 32 to 8 electrodes: AF3, AF4, F3, F4, F7, F8, T7 and T8, as these electrodes may capture better emotion states.

2.2.1.3 Emotion measurement: valence and arousal

As presented in the previous chapter, valence and arousal values are represented in the range [1-9]. From this range we can differentiate three sections:

- Low values: range 1 to 3.
- Neutral values: range 4 to 6.
- High values: range 7 to 9.

In order to this, we divided the representation state in nine different sections, obtaining nine different classes:

- **Happy**: high valence and high arousal (HVHA).
- **Pleased**: high valence and neutral arousal (HVNA).
- **Relaxed**: high valence and low arousal (HVLA).
- **Excited**: neutral valence and high arousal (NVHA).
- **Neutral**: neutral valence and neutral arousal (NVNA).
- **Calm**: neutral valence and low arousal (NVLA).
- **Distressed**: low valence and high arousal (LVHA).
- **Miserable**: low valence and neutral arousal (LVNA).
- **Depressed**: low valence and low arousal (LVLA).

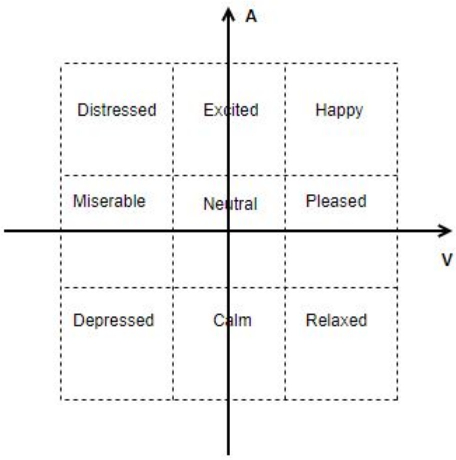

Figure 2.2.1: Graphical representation of the classification space.

Since valence and arousal values are part of the DEAP dataset, we did not have to calculate them.

In our classification methods, we have used two different sets of valence and arousal values: one with the rating values of the participants and another one with the stimuli valence and arousal values.

2.2.1.4 Classification models

There are several emotion classification models, for different purposes and data structures. Due to the nature of our data and how it is labelled we choose both unsupervised machine learning and supervised machine learning models:

- **K-Means**: is a clustering algorithm based on vector quantization. This algorithm aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. The goal of using this algorithm is to check whether the data is enough self-explanatory to form different groups which represent our different emotions.

- **Support-Vector Machines** (SVM): is one of the most used supervised learning algorithms. A support-vector machine constructs a hyperplane or a set of hyperplanes in a high-dimensional space, which can be used for classification, regression or other machine learning tasks. A good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (called functional margin). SVMs use kernel functions $k(x, y)$ to ensure that mappings used by SVMs schemes can be easily computed in terms of the variables in the original space.

- **Random Forest**: are an ensemble learning method for classification, regression and other tasks. It constructs a multitude of decision trees at training time and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

- **Classification Trees**: this method of machine learning uses a decision tree (as a predictive model) to go from observations about an item (represented int he branches) to conclusions about the item's target value (represented in the leaves). In these structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels.

Figure 2.2.2 shows a representation of the input and output of our classifiers, being the input a supervector with the EEG data signal along with the extracted features described in section 2.2.1.1 and the output the emotion (represented as colours) for each observation.

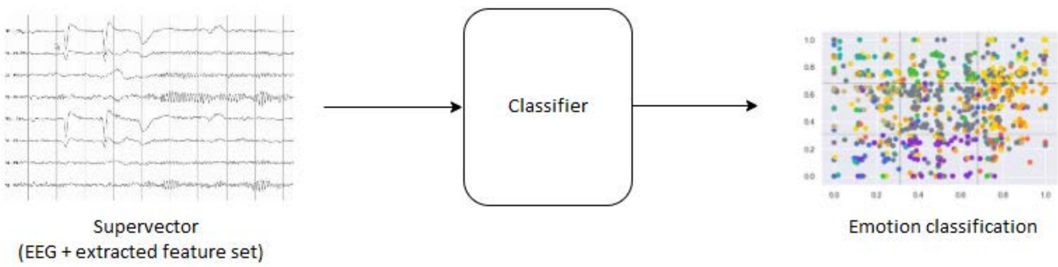

Figure 2.2.2: Classification process schema.

2.2.1.5 Data preprocessing

Although the data we are using is preprocessed, in order to implement the classifiers presented in the previous section, we have to normalise our data and divide in train and test sets.

We normalise the data using `MinMaxScaler`, which essentially shrinks the range to [0,1]:

$$x_{norm} = \frac{x_i - min(x)}{max(x) - min(x)} \tag{2.2.9}$$

Train and test sets are built in a 80-20 proportion using the functions of `sklearn` library.

2.2.2. Stimuli analysis

This section presents the analysis and feature extraction of the musical stimuli used for the experiments in the DEAP dataset. The goal of this study is try to find any relation between musical features and emotions.

We have analysed a total of 40 songs, which are listed in appendix A.

2.2.2.1 Feature extraction

The feature extraction process aims to calculate a numerical representation of music. Music signals are time-varying signals with wider bands of frequency than human vocal sound and many elements. Therefore, we extract features using time-frequency analysis and basic elements related to rhythm and harmony.

To facilitate the analysis, and since that was the length of the stimuli used in the DEAP dataset, we only used the first minute of the songs.

Basic theoretical musical features

These added features are related to the fundamental elements of music: rhythm, dynamics, melody, harmony, tone colour, texture and form. In order to simplify the analysis, and since our set of songs are from different styles and genres, we chose three basic characteristics:

- **Tempo**: related to rhythm. The tempo is the speed of the beat, which can be described by the number of beats per second.

- **Key**: related to harmony. The key of a piece is the group of pitches, or scale, that forms the basis of a music composition. The group features a tonic note and its corresponding chords, also called tonic, which provides a subjective sense of arrival and rest. Notes and chords other than the tonic in a piece create varying degrees of tension, resolved when the tonic note or chord returns.

- **Mode**: the key may be in the major or minor mode. The hallmark that distinguishes major keys from minor is whether the third scale degree is major or minor. It changes the mood of the music.

Time-frequency features

Since music signals are represented as time series signals, we chose some of the most used time-frequency features for music analysis:

- **Statistics**: maximum, minimum, mean and standard deviation values of the musical signal.

- **Zero crossing rate**: is the rate of sign-changes along a signal. It usually has higher values for highly percussive sounds like those in metal and rock. ZCR is defined formally as:

$$zrc = \frac{1}{T-1} \sum_{t=1}^{T-1} 1_{R_{<0}}(s_t s_{t-1}) \tag{2.2.10}$$

  where s is a signal of length T and $1_{R_{<0}}$ is an indicator function.

- **Mean and standard deviation of spectral centroid**: the spectral centroid is a measure used in digital signal processing to characterise a spectrum. It indicates where the "centre of mass" of the spectrum is located. Perceptually, it has a robust connection with the impression ob "brightness" of a sound. It is calculated as the weighted mean of the frequencies present in the signal, determined using a Fourier transform, with their magnitudes as the weights:

$$Centroid = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \tag{2.2.11}$$

where $x(n)$ represents the weighted frequency value, or magnitude, of bin number n, and $f(n)$ represents the centre frequency of that bin.

- **Spectral flux**: is a measure of how quickly the power spectrum of a signal is changing, calculated by computed the power spectrum for one frame against the power spectrum from the previous frame. It is usually calculated as the 2-norm between the two normalised spectra.

$$SF(n) = \frac{\sum_{k=-N/2}^{N/2-1} H(|X(n,k)| - |X(n-1,k)|}{\sum_{k=-N/2}^{N(2-1} |X(n,k)|} \tag{2.2.12}$$

  It can be used to determine the timbre of an audio signal.

- **Mean and standard deviation of Tonnetz**: the Tonnetz is a conceptual lattice diagram representing tonal space. The Tonnetz organises equal-tempered pitch on a conceptual planed according to intervallic relations, favouring perfect fifths, major thirds and minor thirds.

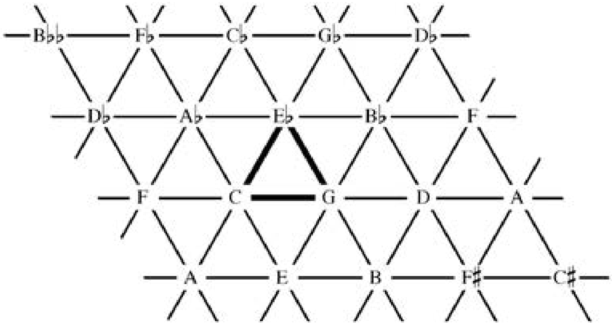

Figure 2.2.3: Triangle Tonnetz.

The nodes represents *pitches* (pitch classes) and the edges are the *consonant intervals* (m3/M6, M3,m6, P4/P5) creating triangles that represents *chords*. In figure 2.2.3, the highlighted triangle represents the chord of C minor (established by the notes C, Eb and G).

Mathematically, we express this multi-level pitch configuration (represented in the Tonal Interval Space by TIVs $T(k)$) as explained in [Bernades et al., 2016]:

$$T(k) = w(k) \sum_{n=0}^{N-1} \bar{c}(n) e^{\frac{-j\pi kn}{N}}, k \in \mathbb{Z} \tag{2.2.13}$$

with

$$\bar{c}(n) = \frac{c(n)}{\sum_{n=0}^{N-1} c(n)} \tag{2.2.14}$$

- **Chromagram**: chroma features are an interesting and powerful representation for music audio in which the entire spectrum is projected onto 12 bins representing the 12 distinct semitones (or chroma) of the musical octave. Since, in music, notes exactly one octave apart are perceived as particularly similar, knowing the distribution of chroma even without the absolute frequency can give useful musical information about the audio, and may even reveal perceived musical similarity that is not apparent in the original spectra. As explained in [Müller and Balke, 2015], given a pitch-based log-frequency spectrogram

$$\gamma_{LF} : \mathbb{Z} \times [0:127] \rightarrow \mathbb{R}_{\geq 0} \tag{2.2.15}$$

defined by

$$\gamma_{LF}(m,p) := \sum_{k \in P(p)} |\mathcal{X}(m,k)|^2 \tag{2.2.16}$$

where we define for each pitch $p \in [0:127]$ the set

$$P(p) := \{k \in [0:K] : F_{pitch}(p-0.5) \leq F_{coef}(k) < F_{pitch}(p+0.5)\} \tag{2.2.17}$$

the chroma representation is derived by summing up all pitch coefficients which belong to the same chroma:

$$\mathcal{C}(m,c) := \sum_{p \in [0:127] | p mod 12 = c} \gamma_{LF}(m,p) \tag{2.2.18}$$

for $c \in [0:11]$.

2.2.2.2 Feature analysis

In order to find some relation between the features outlined in the previous section and the emotion associated with the songs, we will compare these features for each class.

As described in the analysis provided by several other studies in this domain [Eerola, 2012, Jaquet et al., 2012, Le Groux and Verschure, 2010], we can relate sets of features for particular emotions:

- **Tempo**: fast tempo is usually associated with high arousal, while a slower tempo is related to low arousal values.

- **Key mode**: major mode is associated with high valence values and minor mode is associated with low valence values.

- **Spectral centroid deviation**: a higher deviation is related to higher arousal values.

- **Pitch level variation**: its effects are more "ambiguous" and "complex". However, it seems that increasing pitch levels are related to faster tempos, which lead us to higher arousal values.

- **Zero crossing rate**: as explained in the previous section, zero crossing rate is related to the pitch. Higher values mean more strident sounds (the ones we can find in rock or metal music which shows usually high arousal values and neutral to low valence values) while lower values are associated with "rounder" sounds (the ones we can find in jazz music which shows lower arousal values and neutral to high valence values).

Furthermore, we will use the chromagrams of each song to check and verify the similarity between songs with the same emotion label.

2.3. Software libraries

All the methods and algorithms described in this chapter where performed using `Python 3` and specific libraries for signal and time-series data analysis:

- `Pandas`: data manipulation as DataFrames.
- `Scipy` and `Numpy`: statistics.
- `Sklearn`: machine learning algorithms.
- `cPickle`: DEAP dataset files reading.
- `Librosa`: time-frequency music analysis.

3. Results

This chapter contains the results obtained with all the developed experiments and analysis, both for emotion classification and stimuli analysis.

We need to consider and emphasise that we are using **nine** different classes in order to correctly interpret and understand the results.

3.1. Emotion classification

We carried out two different classification studies:

- **Participant emotion recognition**: we use the valence and arousal values provided by the participants during the experiments. Since we are using the labels given by the participants, we try to recognise the emotion that the participants are feeling using biosignals data, without taking account of the stimuli they are being exposed to.

- **Video emotion recognition**: we use the valence and arousal mean values registered in the file `Videos.csv`. In this case, we try to correctly classify the emotion associated with the stimuli.

For both recognition process we implemented all the algorithms explained in the previous chapter, along with additional experiments:

- **Feature selection via grid search with Decision Tree**: due to the results obtained with the Decision Tree algorithm we decided to perform the feature selection using this method too and compare the results.

- **Comparison between classification using EEG data alone and using extracted features alone**: in order to see how explanatory our extracted features can be for emotion recognition.

3.1.1. Participant emotion recognition

3.1.1.1 Feature selection

As explained in the previous chapter, we perform feature selection using a Grid Search Cross-Validation method with Random Forest as estimator. We select a total of 7692 features out of 8080, discarding only around 200 features which are points from the EEG data signal.

Furthermore, we use Decision Tree as estimator too, getting a different and smaller set of selected features. Using Grid Search Cross-Validation with Decision Tree as estimator we select 315 features out of 8080. We select only 300 points from the EEG data signal and discard EEG mean and complexity values from the extracted features.

3.1.1.2 Algorithm parameters

The parameters for the different implemented algorithms are listed below:

- **Support Vector Classification (SVC)**:
  - C = 1000
  - cache_size = 200
  - decision_function_shape = ovo
  - gamma = 0.5
  - kernel = rbf
  - tol = 0.001
- **Random Forest (RF):**
  - n_estimators = 601
  - random_state = 123456
- **Decision Tree (DT):**
  - criterion = entropy
  - random_state = 0
  - presort = True

These values were estimated running a few test, both manual and hyper parameter tuning.

3.1.1.3 Classification Results

Tables 3.1.1 and 3.1.2 show the results obtained with the set of selected features using Random Forest and Decision Tree as estimators respectively.

We can see that the selected feature set via Decision Tree performs better than the Random Forest one, except for SVC. This might be due to the need of more data for SVC to classify properly with many classes.

In both cases, K-Means does not obtain good results (below 10%). This was somehow expected since it is an unsupervised learning method.

Classification Tree seems to be the best classifier in both sets.

Algorithm	Accuracy
K-Means	9.131%
SVC	59.765%
RF	52.197%
DT	82.666%

Table 3.1.1: Participants emotion classification results with Random Forest feature selection.

Algorithm	Accuracy
K-Means	11.132%
SVC	46.484%
RF	65.381%
DT	85.937%

Table 3.1.2: Participants emotion classification results with Decision Tree feature selection.

Graph results are shown in appendix B.

3.1.1.4 Dependence on the sampling rate

As explained in section 2.2.1.2, we create another set of features changing the sampling rate of the EEG signal, selecting one out of n points. In our case, we select from 1 out of 2 to 1 out of 64, which would correspond to a 2Hz frequency.

Table 3.1.3 shows the results for the different classifiers with the selected point sets.

Points (1 out of n)	**SVC Accuracy**	**RF Accuracy**	**DT Accuracy**
2	58.105%	53.906%	82.129%
8	48.975%	58.691%	86.963%
16	48.829%	62.061%	88.672%
32	47.705%	65.967%	89.941%
64	42.089%	71.875%	91.259%

Table 3.1.3: Sampling rate experiments participant emotion classification results selecting 1 point out of n.

Again, the best classifier is Decision Tree. However, the accuracy of Random Forest seems to improve as the EEG data signal get reduced, obtaining an almost 20% accuracy enhancement.

On the other side, SVC get worse as the set get reduced. It shows a similar behaviour to the results with the selected features sets of Random Forest and Decision Tree estimators, where the smaller the set is, the worse it performs).

3.1.1.5 EEG vs Extracted features

For the purpose of assessing how explanatory are the extracted features, we tried to classify the data using EEG data signals alone and extracted features alone. This would allow us to know if the extracted features are adding valuable information to the classifier or, on the contrary, adding noise and useless data.

The extracted features used are mean and kurtosis EEG signal values, Hjorth features, absolute and relative spectral power values in β and γ bands and respiration and temperature statistical values (maximum, minimum, mean and standard deviation VALUES).

The results are shown in table 3.1.4.

	SVC Accuracy	RF Accuracy	DT Accuracy
EEG alone	56.054%	50.049%	25.732%
Extracted features	29.345%	98.779%	95.459%

Table 3.1.4: Emotion classification results for EEG data alone and extracted features data alone.

EEG data signal alone seems to provide similar results for SVC and Random Forest classifiers with feature selection via Random Forest estimator. However, the results of Decision Tree classifier are far worse.

On the other side, extracted features set shows a very nice performance for both Random Forest and Decision Tree classifiers, in contrast to SVC results.

We can observe that the results using only the extracted features are better than use EEG alone as input for the classifiers. This suggest that the extracted feature set is quite informative. Moreover, these results are consistent with the ones obtained in the previous section, where the more reduced the input set was, the better the classifier performed. This might be due to the EEG signal nature, which is a very noisy and complex signal.

3.1.2. Video emotion recognition

3.1.2.1 Feature selection

Like the previous classification problem, we performed feature selection using a Grid Search Cross-Validation method with Random Forest and Decision Tree as estimators. We selected a total of 510 features out of 8080 with Random Forest estimator and 315 with Decision Tree. In both cases, the selected feature sets included a few points with some of the added features like Gamma bandpower, respiration data or temperature data.

3.1.2.2 Algorithm parameters

The parameters for the different implemented algorithms are listed below:

- **SVC**:
  - C = 1000

- cache_size = 200
- decision_function_shape = ovo
- gamma = 0.9
- kernel = rbf
- tol = 0.001

- **RF:**
  - n_estimators = 606
  - random_state = 123456
- **DT:**
  - criterion = entropy
  - random_state = 0
  - presort = True

These values were estimated running a few test, both manual and hyper parameter tuning.

3.1.2.3 Classification Results

Tables 3.1.5 and 3.1.6 show the results obtained with the set of selected features using Random Forest and Decision Tree as estimators, respectively.

We obtained very similar results to the ones for the participants emotion recognition, being the main difference that Decision Tree does not perform well with the Random Forest feature selected set. Again, the selected feature set via Decision Tree performs better than the Random Forest one, except for SVC and K-Means, which get slightly worse in this case.

In both cases, K-Means still does not obtain good results.

Another difference between the results of the participants emotions classification and video emotion classification is with the Random Forest classifier, which is quite better with the Decision Tree feature selected set in the last one. This indicates, once again, that the Decision Tree feature set is better than the Random Forest one.

Lastly, as expected, Classification Tree gets the best result.

Algorithm	Accuracy
K-Means	11.377
SVM	53.515%
RF	53.027%
DT	30.322%

Table 3.1.5: Video emotion classification results with Random Forest feature selection.

Algorithm	Accuracy
K-Means	11.230
SVM	36.377%
RF	79.882%
DT	95.703%

Table 3.1.6: Video emotion classification results with Decision Tree feature selection.

Graph results are shown in appendix B.

3.1.2.4 Dependence on the sampling rate

Table 3.1.7 shows the results for the different classifiers with the selected point sets.

Points (1 out of n)	SVC Accuracy	RF Accuracy	DT Accuracy
2	60.009%	56.933%	87.060%
8	51.367%	57.812%	85.302%
16	55.371%	58.251%	91.406%
32	51.806%	60.644%	92.382%
64	44.580%	64.794%	94.531%

Table 3.1.7: Sampling rate experiments video emotion classification results selecting 1 point out of n

We observe the same performance as the participants emotion classification, getting slightly worse results for Random Forest classifier and better for both SVC and Decision Tree classifiers.

3.1.1.5 EEG vs Extracted features

Regarding how explanatory our selected features can be for video emotion classification, the results are shown in table 3.1.8.

Again, we obtain similar results and the same behaviour as the participants emotion classification, being these ones slightly worse than the obtained with video emotion.

	SVC Accuracy	RF Accuracy	DT Accuracy
EEG	60.644%	56.835%	31.933%
Extracted features	32.812%	99.804%	96.728%

Table 3.1.8: Emotion classification results for EEG data alone and extracted features data alone.

3.2. Stimuli analysis

First, we grouped the songs by classes having the following sets of songs:

- Class 1 (Happy): one song (experiment 5).
- Class 2 (Pleased): eleven songs (experiments 1, 3, 4, 7, 8, 9, 11, 13, 14, 18 and 19).
- Class 3 (Relaxed): none.
- Class 4 (Excited): three songs (experiments 2, 10 and 36).
- Class 5 (Neutral): twelve songs (experiments 6, 15, 20, 21, 22, 25, 31, 32, 33, 34, 35 and 40).
- Class 6 (Calm): five songs (experiments 12, 16, 17, 26 and 27).
- Class 7 (Distressed): none.
- Class 8 (Miserable): six songs (experiments 23, 29, 30, 37, 38 and 39).
- Class 9 (Depressed): two songs (experiments 24 and 28).

As we can see, we cannot characterise some of the classes due to the lack of examples in them, having none (classes 3 and 7) or just a few (classes 1 and 9).

Table 3.2.1 shows the mean results for each class:

Class	Max T	Min T	Mean T	Major	Minor	ZCR	SCD
Happy	130	130	130	1	0	162404	749.861
Pleased	176	77	121	7	4	139825	915.298
Excited	174	118	148	3	0	146324	865.815
Neutral	171	63	105	7	6	143070	736.459
Calm	130	80	111	3	2	88587	837.152
Miserable	185	74	140	3	3	135505	668.011
Depressed	120	90	105	2	0	77150	903.357

Table 3.2.1: Features mean values for each class: max tempo, min tempo, mean tempo, major modes count, minor modes count, zero crossing rate and spectral centroid deviation.

We can observe several outstanding features:

- The max tempo value is related to the *Miserable* class, which is a class of low valence and neutral arousal. Usually, this values are associated with lower tempo values. Although it could be an outlier, the mean tempo value is the second faster mean tempo of the results, which is very interesting.

- In relation to major and minor modes, we observe that *Neutral* has even number of both modes, same as *Miserable*. Major modes are associated with high valence values, so the results for *Depressed* and *Miserable* are also interesting.

ZRC and SCD values are close to what we expected:

- Higher ZRC for classes with higher arousal.

- Higher SCD for classes with higher arousal.

Regarding the chromagrams, whose graphs are in appendix C, we find the following characteristics:

- Class *Please* chromagrams usually start with a few seconds of silence.

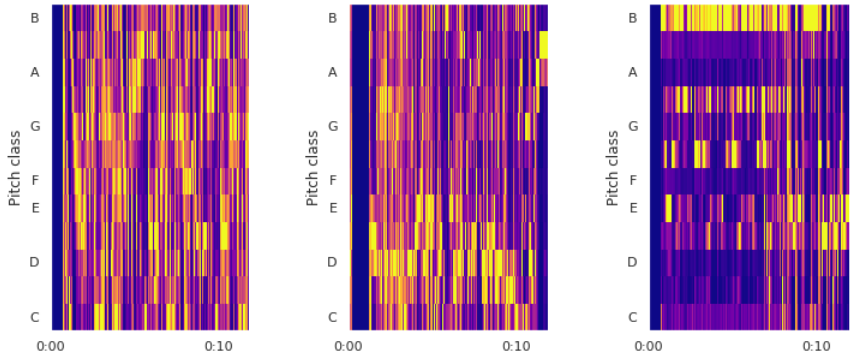

Figure 3.2.1: Songs number 3 (right), 4 (centre) and 8 (left) starts, where we appreciate silence at the begining of the songs.

- Class *Calm* have three songs (numbers 16, 17 and 26) with similar patterns, having two main lines.

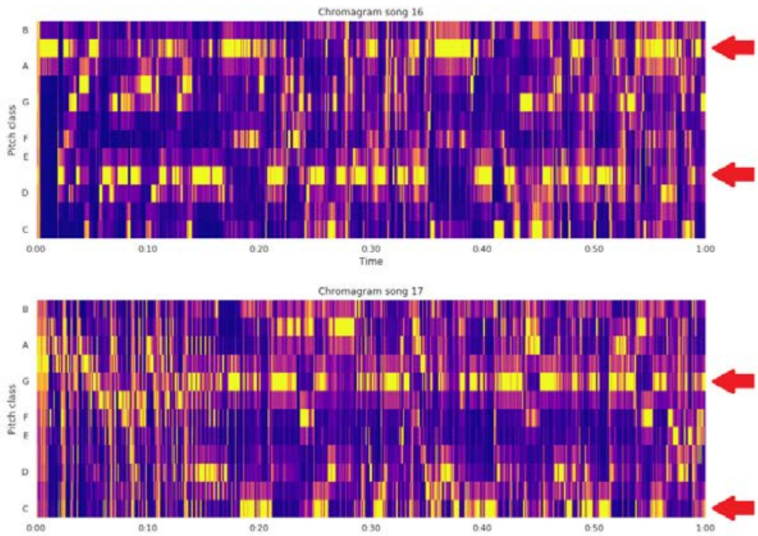

Figure 3.2.2: Songs number 16 and 17. We can observe two main lines which match tones with fundamental-fifth relation (Eb-Bb and C-G respectively).

- Classes *Depressed* and *Excited* have patterns that resembles “arpeggios” or “scales”, with different grade of density due to the tempo.

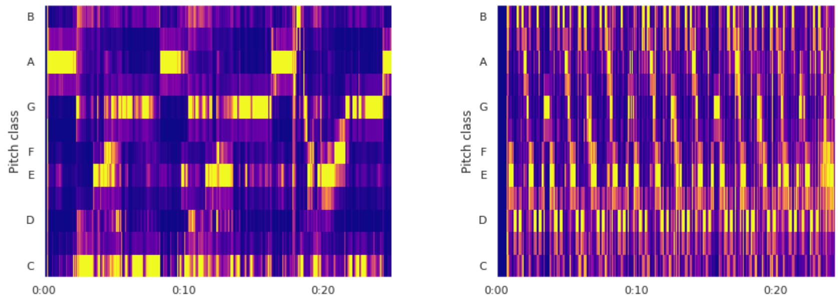

Figure 3.2.3: *Arpeggios* sections from songs 28 (right) and 36 (left), classes *Depressed* and *Excited* respectively.

Classification results

We tried to classify the songs according to their classes using the features that we extracted for each song. Again, we performed feature selection via Decision Tree estimator and K-Means and Random Forest as classifiers. The results are shown in the table 3.2.2.

Algorithm	Accuracy
K-Means	12.5%
RF	37.5%

Table 3.2.2: Songs emotion classification results with Decision Tree feature selection.

As we expected, the classification results are not good. Graphs results are in appendix C.

4. Conclusions

4.1. Discussion

This project had two main objectives:

- The analysis and classification of emotions using biological signals and musical stimuli.
- The analysis of the relation between the musical stimuli and the emotions.

Results regarding the first goal are very satisfactory, as we have achieved above 90% accuracy with the Decision Tree classifier, reaching the following conclusions:

- Due to the representation of our data in a bi-dimensional space, Random Forest and Decision Tree are the best classifiers for our problems (participant emotion classification and video emotion classification).

- Regarding the results changing sample rate (shown in tables 3.1.3 and 3.1.7), apparently a high time resolution in EEG it is not necessary to classify emotion via musical stimuli. This may have sense since noticeable or relevant changes in music do not need to happen in small short time periods (e.g. milliseconds), but in larger ones (half a second or even several seconds).

- The results shown in tables 3.1.4 and 3.1.8 suggest that the extracted features from the biological signals seem to be very self-explanatory and informative of the data in relation to emotions. This suggest that our set of extracted features defines properly the nature of the data.

- Although our set of features is very informative, unsupervised learning does not perform well. This suggests that the data is still very complex, heterogeneous and disperse, not having clear "centroids" or a clear distribution.

As for the stimuli analysis, there are some interesting outputs, regarding tempo and key mode features.

As explained in section 2.2.2, tempo is a feature associated to arousal while key mode is associated to valence. However, in table 3.2.1, classes *Miserable* and *Depressed*, which have low valence values and neutral-low arousal valence, do not show that relation. The same effect happens to class *Pleased*, with neutral arousal values and high valence values.

This suggest that, as we mention in the introduction, music is a very complex signal and we cannot infer what kind of emotion it recreates only taking into account two factors (valence and arousal). This conclusion is supported by the results we obtain in table 3.2.2 and the different chromagrams in C, which are very different even having the same class. That it because different music genres and styles can cause the same emotion. However, we cannot generalise any results since our dataset is very small.

As is mention in other studies [Jaquet et al., 2012], maybe we need to focus on study the inter-relation between several factors in order to established a proper relation between musical features and emotion recognition.

In conclusion, with this project we achieved our goal of characterise biological signals for emotion recognition with musical stimuli and find interesting outcomes regarding music features and emotions.

4.2. Future work

- Future work could address the analysis of the resting state dynamics before and after the stimuli. This could provide very interesting information about how the singularity of the individual EEG signal can affect emotions.

- It would be highly informative to perform the stimuli analysis using a large dataset with more songs and music examples in order to obtain more general results and being able to draw conclusions more accurately.

- EEG sonification tools [Sanyal et al., 2019] may contribute to establish further links between EEG and music stimuli with regard to emotion classification.

Bibliography

[Aungsakul et al., 2012] Aungsakul, S., Phinyomark, A., Phukpattaranont, P., and Limsakul, C. (2012). Evaluating feature extraction methods of electrooculography (eog) signal for human-computer interface. *Proceda Engineering*, 32:246–252.

[Baydogan and Runger, 2015] Baydogan, M. G. and Runger, G. (2015). Learning a symbolic representation for multivariate time series classification. *Data Mining and Knowledge Discovery*, 29(2):400–422.

[Bernades et al., 2016] Bernades, G., Cocharro, D., Caetano, M., Guedes, C., and Davies, M. (2016). A multi-level tonal interval space for modelling pitch relatedness and musical consonance. *Journal of New Music Research*, 45(4):281–294.

[Bradley and Lang, 1994] Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49 – 59.

[Dong et al., 2018] Dong, Y., Yang, X., Zhao, X., and Li, J. (2018). Bidirectional convolutional recurrent sparse network (bcrsn): An eficient model for music emotion recognition. *IEEE Transactions on Multimedia*.

[Eerola, 2012] Eerola, T. (2012). Modeling listeners' emotional response to music. *Topics in Cognitive Science*, 4:607–624.

[Ehrlich et al., 2019] Ehrlich, S. K., Agres, K. R., Guan, C., and Cheng, G. (2019). A closed-loop, music-based brain-computer interface for emotion mediation. *PLoS one*, 14(3).

[Hamada et al., 2018] Hamada, M., Zaidan, B. B., and Zaidan, A. A. (2018). A systematic review for human eeg brain signals based emotion caassification, feature extraction, barin condition, group comparison. *Journal os Medical Systems*, pages 42–162.

[Jäncke, 2008] Jäncke, L. (2008). Music, memory and emotion. *Journal of Biology*, 7(6):21.

[Jaquet et al., 2012] Jaquet, L., Danuser, B., and Gomez, P. (2012). Music and felt emotions: how systematic pitch level variations affect the experience of pleasantness and arousal. *Psychology os Music*, 1:51–70.

[Jatupaiboon et al., 2013] Jatupaiboon, N., Pan-ngum, S., and Israsena, P. (2013). Real-time eeg-based happines detection system. *The Scientific World Journal*, pages 1–12.

[Jenke et al., 2014] Jenke, R., Peer, A., and Buss, M. (2014). Feature extraction and selection for emotion recognition from eeg. *IEEE Transactions On Affective Computing*, 5(3):327–339.

[Jirayucharoensak et al., 2014] Jirayucharoensak, S., Pan-Ngum, S., and Israsena, P. (2014). Eeg-based emotion recognition using deep learning network with principal component based covariate shift adaptation. *The Scientific World Journal*, pages 1–10.

[Koelstra et al., 2012] Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., and Ebrahimi, T. (2012). Deap: A database for emotion analysis ;using physiological signals. *IEEE Transactions On Affective Computing*, 3(1):18–31.

[Le Groux and Verschure, 2010] Le Groux, S. and Verschure, P. F. (2010). Emotional responses to the perceptual dimensions of timbre: A pilot study using physically informed sound synthesis.

[Müller and Balke, 2015] Müller, M. and Balke, S. (2015). *Short-Time Fourier Transform and Chroma Features.* Friedrich-Alexander-Universität Erlangen-Nürnberg.

[Posner et al., 2005] Posner, J., Russel, J., and Peterson, B. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev Psuchopathol*, 17(3):715–734.

[Pourahmadi and Noorbaloochi, 2016] Pourahmadi, M. and Noorbaloochi, S. (2016). Multivariate time series analysis of neuroscience data: some challenges and opportunities. *Current opinion in neurobiology*, 37:12–15.

[Purwins et al., 2008a] Purwins, H., Grachten, M., Herrera, P., Hazan, A., Marxer, R., and Serra, X. (2008a). Computational models of music perception and cognition i: The perceptual and cognitive processing chain. *Physics Of Life Reviews*, 5:151–168.

[Purwins et al., 2008b] Purwins, H., Grachten, M., Herrera, P., Hazan, A., Marxer, R., and Serra, X. (2008b). Computational models of music perception and cognition ii: Domain-specific music processing. *Physics Of Life Reviews*, 5:169–182.

[Rozado et al., 2012a] Rozado, D., Agustin, J. S., Rodriguez, F. B., and Varona, P. (2012a). Gliding and saccadic gaze gesture recognition in real time. *ACM Trans. Interact. Intell. Syst.*, 1(2):10:1–10:27.

[Rozado et al., 2010] Rozado, D., Rodriguez, F. B., and Varona, P. (2010). Optimizing hierarchical temporal memory for multivariable time series. In Diamantaras, K., Duch, W., and Iliadis, L. S., editors, *Artificial Neural Networks – ICANN 2010*, pages 506–518, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Rozado et al., 2012b] Rozado, D., Rodriguez, F. B., and Varona, P. (2012b). Extending the bioinspired hierarchical temporal memory paradigm for sign language recognition. *Neurocomputing*, 79:75 – 86.

[Salimpoor et al., 2015] Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., and McIntosh, A. R. (2015). Predictions and the brain: how musical sounds become rewarding. *Trends in Cognitive Sciences*, 19(2):86–91.

[Sanyal et al., 2019] Sanyal, S., Nag, S., Banerjee, A., Sengupta, R., and Ghosh, D. (2019). Music of brain and music on brain: a novel eeg sonification approach. *Cognitive Neurodynamics*, 13(1):13–31.

[Shen et al., 2019] Shen, J., Tao, M., Qu, Q., Tao, D., and Rui, Y. (2019). Toward efficient indexing structure for scalable content-based music retrieval. *Multimedia systems*, pages 1–15.

[Smith, 2011] Smith, J. O. (2011). *Spectral Audio Signal Processing*. `http://-ccrma.stanford.edu/~jos/sasp/`.

[Song et al., 2012] Song, Y., Dixon, S., and Pearce, M. (2012). Lecture in evaluation of musical features for emotion classification.

[Valenzi et al., 2014] Valenzi, S., Islam, T., Jurica, P., and Cichocki, A. (2014). Individual classification of emotions using eeg. *Journal of Biomedical Science and Engineering*, 7:604–620.

[Vallat, 2018] Vallat, R. (2018). Compute the average bandpower of an eeg signal.

[Varona and Rabinovich, 2016] Varona, P. and Rabinovich, M. I. (2016). Hierarchical dynamics of informational patterns and decision-making. *Proceedings of the Royal Society B*, 283(1832):20160475.

[Zheng and Lu, 2015] Zheng, W.-L. and Lu, B.-L. (2015). Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions On Autonomous Mental Development*, 7(3):162–175.

A. List of songs

- Jungle Drum by Emiliana Torrini
- Scotty Doesn't know by Lustra
- Blame It On The Boogie by Jackson 5
- Love Shack by The B52's
- Song 2 by Blur
- First Date by Blink182
- Satisfaction by Benny Benassi
- Fuck you by Lily Allen
- I Want To Break Free by Queen
- Bombtrack by Rage Against The Machine
- Say Hey (I Love You) by Michael Franti and Spearhead
- Miniature Birds by Bright Eyes
- I'm Yours by Jason Mraz
- Butterfly Nets by Bishop Allen
- Darkest Things bu The Submarines
- Moon Safari by Air
- What A Wonderful World bu Louis Armstrong
- Me gustas tú by Manu Chao
- Love Story by Taylor Swift
- Gloomy Sunday by Diamanda Galas
- Normal by Porcupine Tree
- How To Fight Loneliness by Wilco
- Goodbye My Lover by James Blunt
- Goodbye My Almost Lover by A Fine Frenzy
- The Weight Of My Words by Kings Of Convenience

- Rain by Madonna
- Breathe Me by Sia
- Hurt by Christina Aguilera
- May It Be by Enya
- The One I Once Was by Mortemia
- The Beautiful People by Marilyn Manson
- Bastard Set Of Dreams by Dead To Fall
- Hardcore State of Mind by DJ Paul Elstak
- Procrastination On The Empty Vessel by Napalm Death
- Refuse Resist by Sepultura
- Scorched Earth Erotica by Cradle Of Filth
- Carving A Giant by Gorgoroth
- My Funeral by Dark Funeral
- My Apocalypse by Arch Enemy

B. Classification results graphs

This appendix contains the graphical results of the classification algorithms.

In this graphics the space is divided in nine sections, which represent each emotion “space” in terms of valence and arousal values. The emotions are represented using a specific colour for each one:

- Happy: yellow.
- Pleased: orange.
- Relaxed: red.
- Excited: green.
- Neutral: grey.
- Calm: purple.
- Distressed: aquamarine.
- Miserable: olive.
- Depressed: blue.

Each point represent the valence and arousal values of each observation and the colour assigned to them the class associated according to the classifier.

The next figures show the perfect classification results, the original labels, for each recognition task:

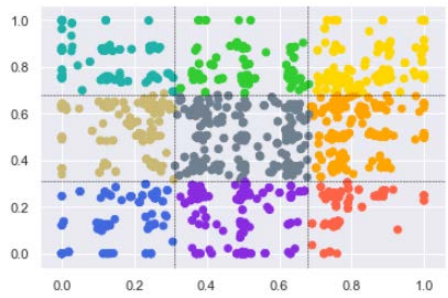

Figure B.0.1: Participants emotion recognition perfect classification (using original assigned labels).

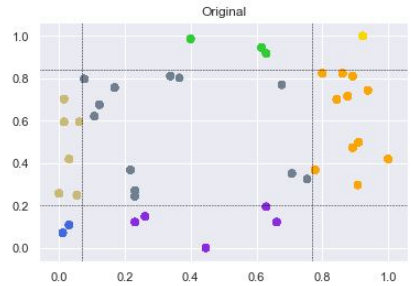

Figure B.0.2: Video emotion recognition perfect classification (using original assigned labels).

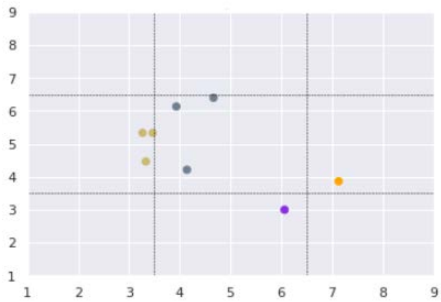

Figure B.0.3: Stimuli recognition perfect classification (using original assigned labels).

We observe that points in the same section have the same colour, since each section is related to one emotion.

B.1. Participants emotion recognition

The graphs are consistent with the results shown in tables 3.1.1 and 3.1.2, where RF and DR performed the best and K-Means the worst. We can highlight a few aspects of these graphs:

- RF seems to classify several observations as *Neutral* in terms of emotion, both with RF and DT feature selection.

- K-Means with RF feature selection only use three classes (*Happy*, *Calm* and *Depressed*).

- K-Means with DT feature selection tends to classify as *Distressed* and *Excited*.

B.1.1. Random Forest estimator

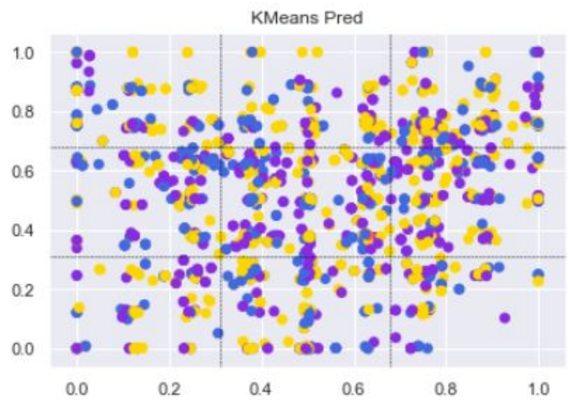

Figure B.1.1: Participants emotion classification K-Means graph results with Random Forest feature selection.

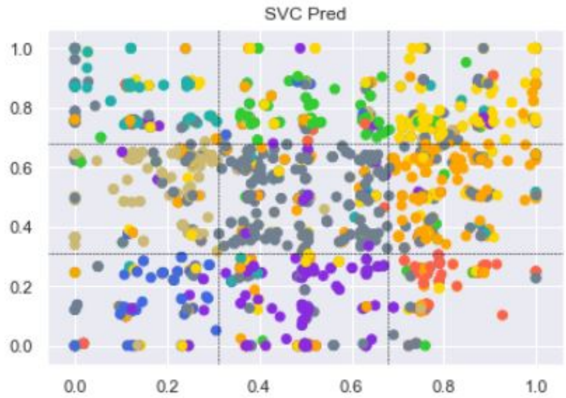

Figure B.1.2: Participants emotion classification with SVC graph results with Random Forest feature selection.

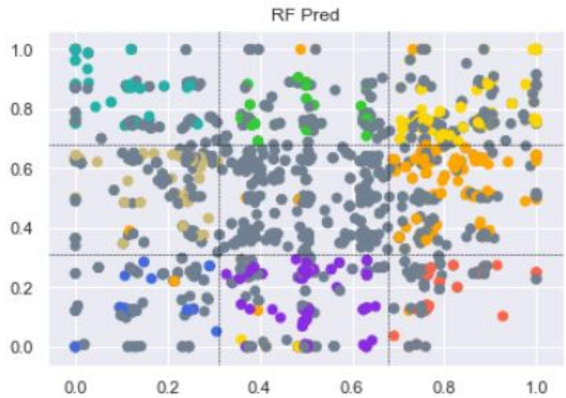

Figure B.1.3: Participants emotion classification with RF graph results with Random Forest feature selection.

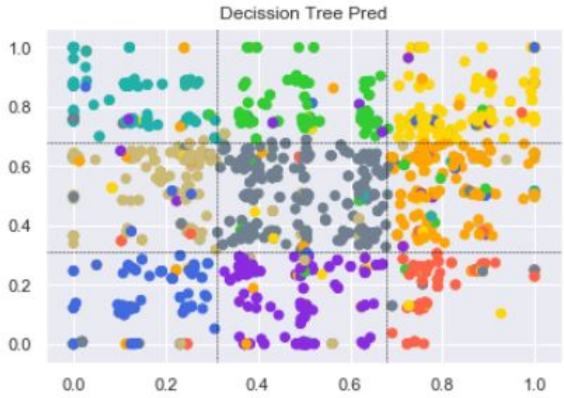

Figure B.1.4: Participants emotion classification with CT graph results with Random Forest feature selection.

B.1.2. Decision Tree estimator

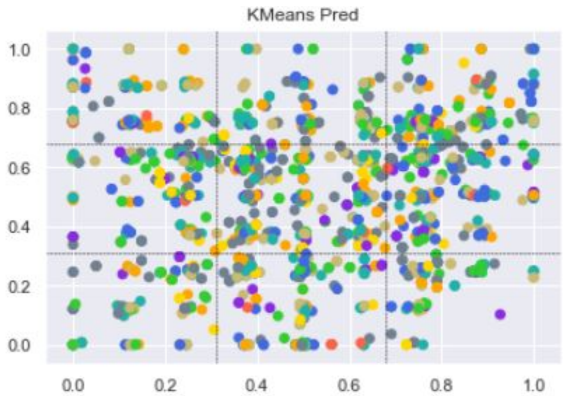

Figure B.1.5: Participants emotion classification K-Means graph results with Decision Tree feature selection.

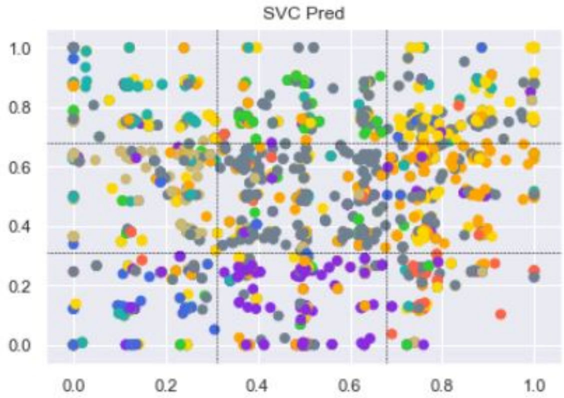

Figure B.1.6: Participants emotion classification with SVC graph results with Decision Tree feature selection.

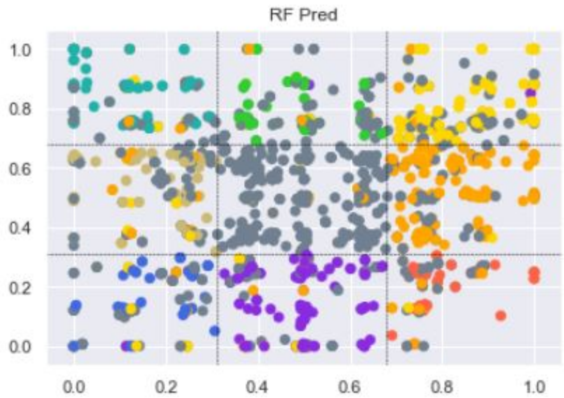

Figure B.1.7: Participants emotion classification with RF graph results with Decision Tree feature selection.

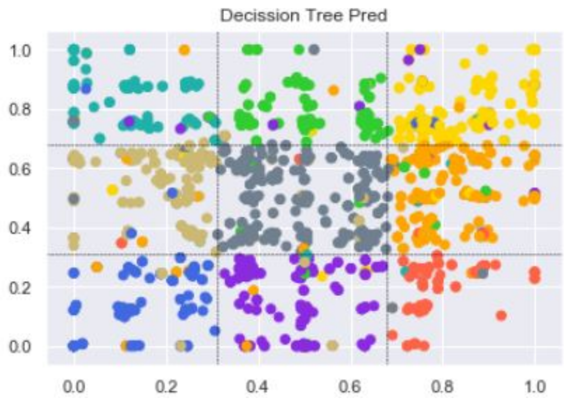

Figure B.1.8: Participants emotion classification with CT graph results with Decision Tree feature selection.

B.2. Video emotion recognition

The graphs are consistent with the results shown in tables 3.1.5 and 3.1.6. In this case, SVC seems to have a tendency to classify several observations as *Neutral* when using DT estimator.

B.2.1. Random Forest estimator

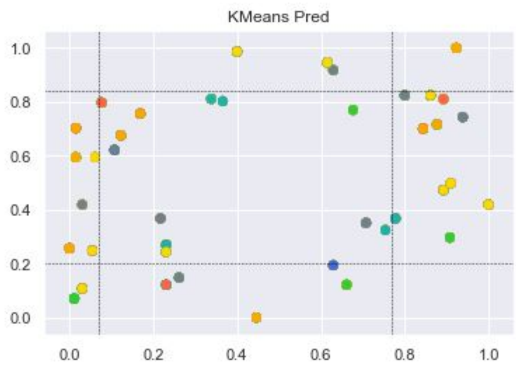

Figure B.2.1: Video emotion classification K-Means graph results with Random Forest feature selection.

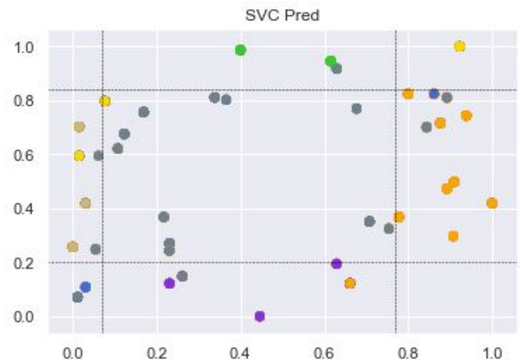

Figure B.2.2: Video emotion classification with SVC graph results with Random Forest feature selection.

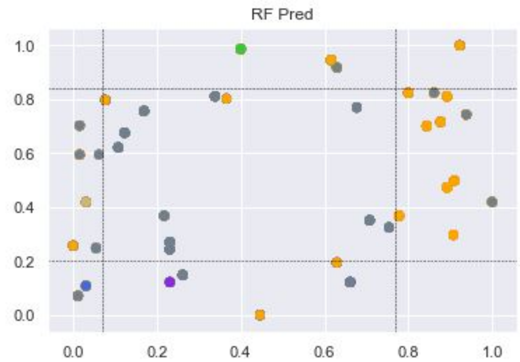

Figure B.2.3: Video emotion classification with RF graph results with Random Forest feature selection.

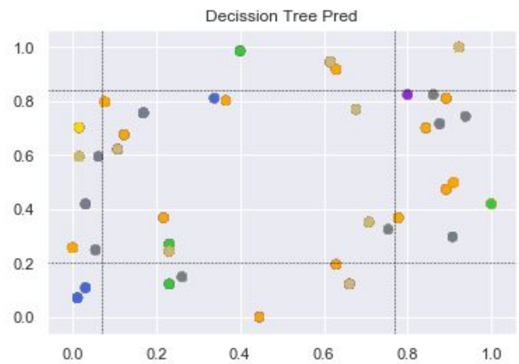

Figure B.2.4: Video emotion classification with CT graph results with Random Forest feature selection.

B.2.2. Decision Tree estimator

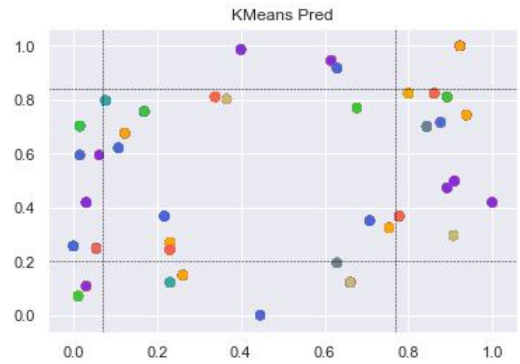

Figure B.2.5: Video emotion classification K-Means graph results with Decision Tree feature selection.

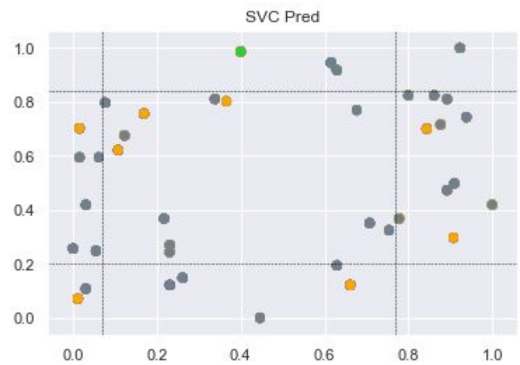

Figure B.2.6: Video emotion classification with SVC graph results with Decision Tree feature selection.

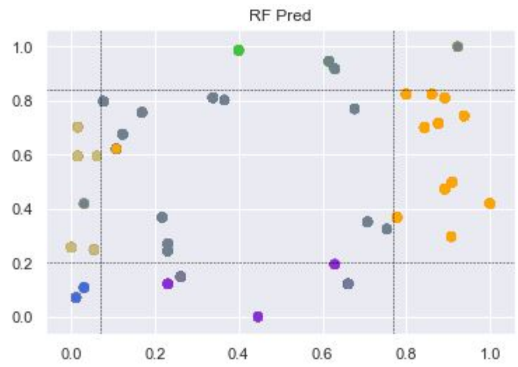

Figure B.2.7: Video emotion classification with RF graph results with Decision Tree feature selection.

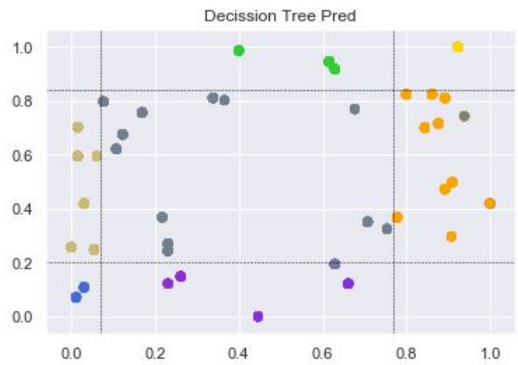

Figure B.2.8: Video emotion classification with CT graph results with Decision Tree feature selection.

B.3. Stimuli classification

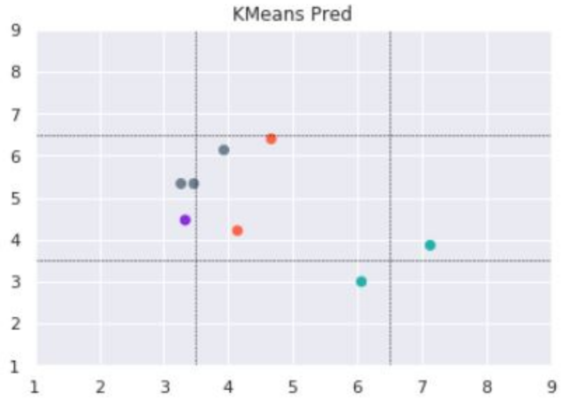

Figure B.3.1: Stimuli emotion classification with KM classifier.

Figure B.3.2: Stimuli emotion classification with RF graph results.

C. Chromagrams

As explained in section 2.2.2, chromagrams are a representation for music audio where we can see the distribution of each note. This allow us to find similarities between different songs with the same label and established possible "features" or "schemes".

C.1. Class happy

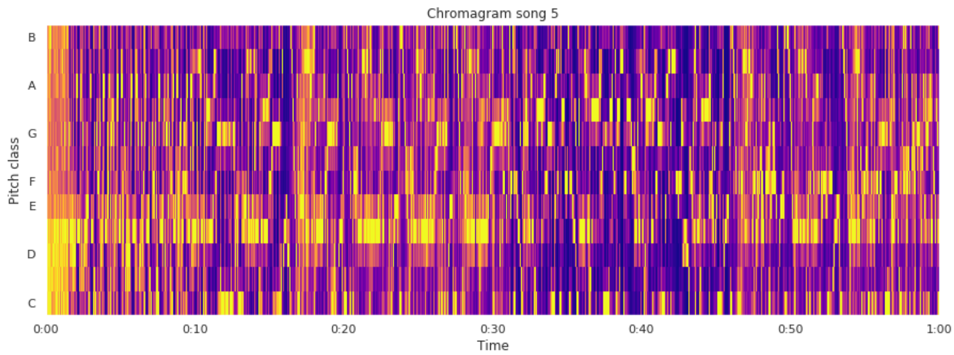

Figure C.1.1: Chromagram song 5.

C.2. Class pleased

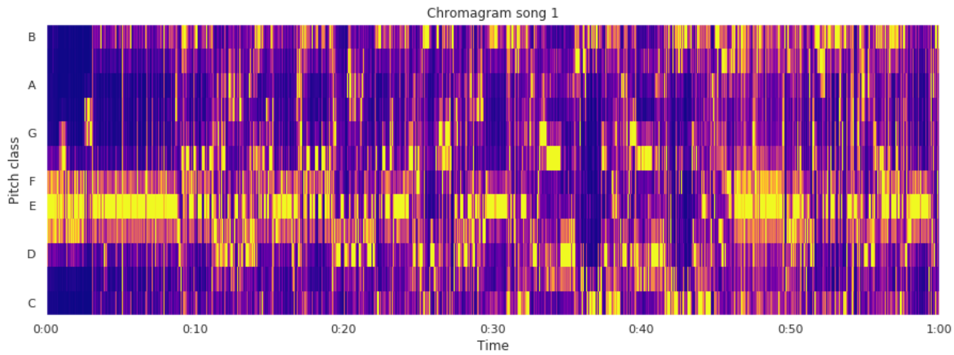

Figure C.2.1: Chromagram song 1.

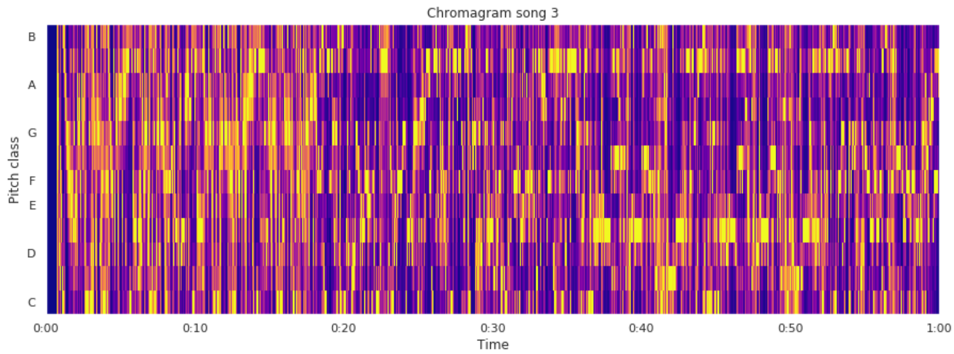

Figure C.2.2: Chromagram song 3.

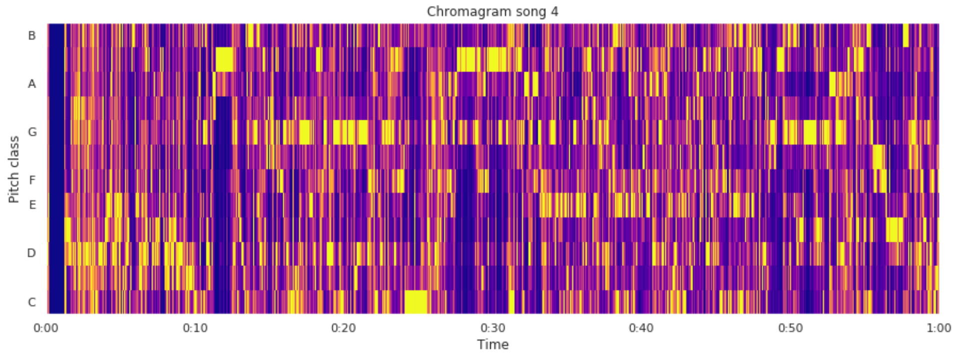

Figure C.2.3: Chromagram song 4.

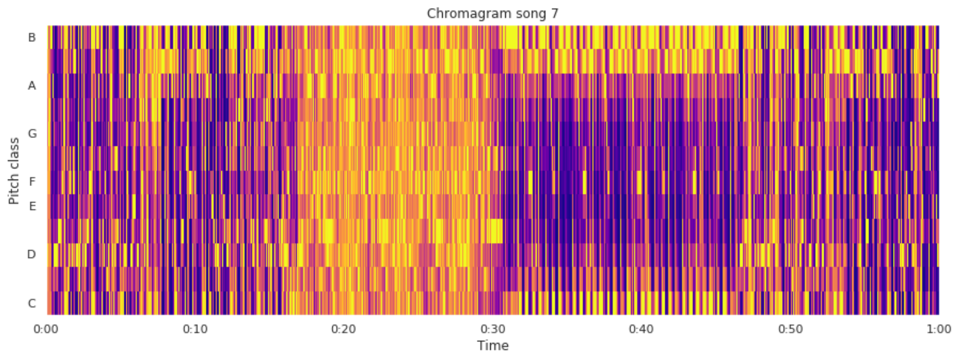

Figure C.2.4: Chromagram song 7.

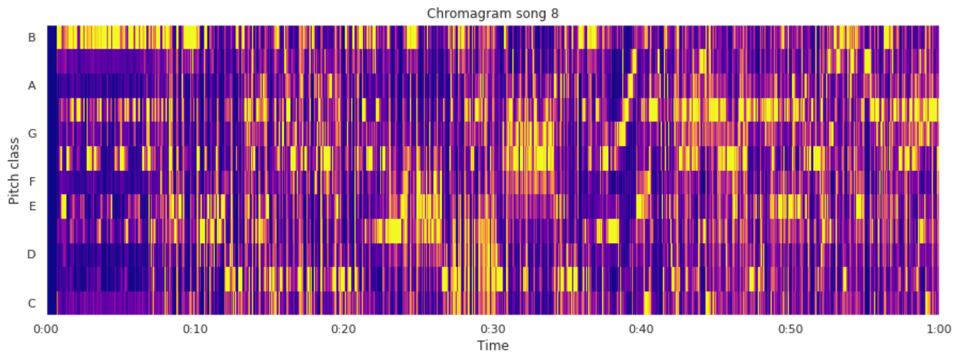

Figure C.2.5: Chromagram song 8.

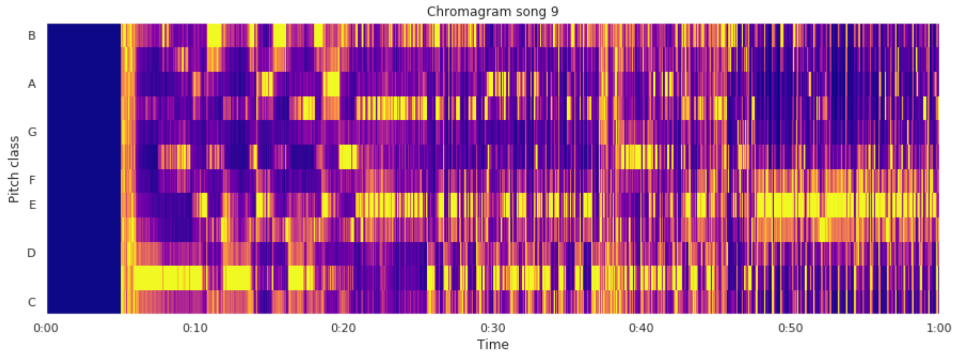

Figure C.2.6: Chromagram song 9.

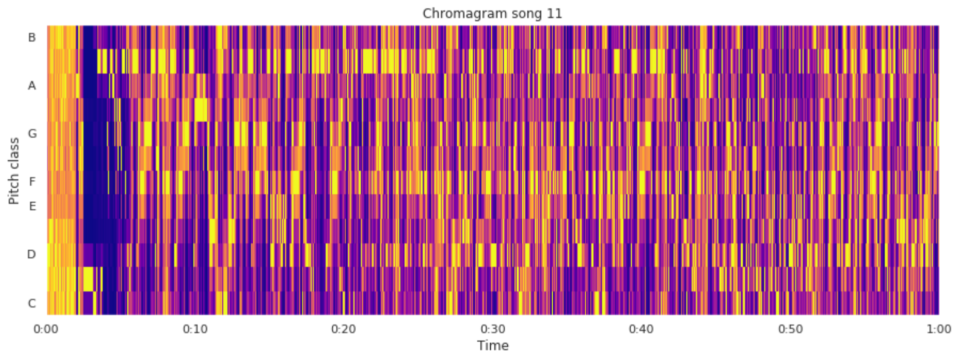

Figure C.2.7: Chromagram song 11.

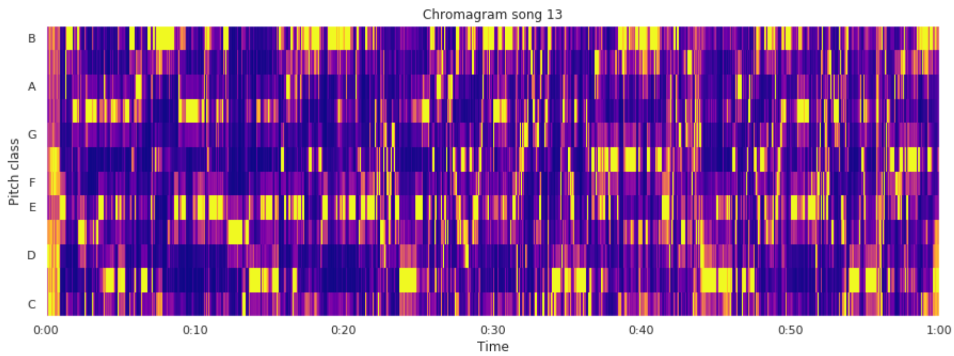

Figure C.2.8: Chromagram song 13.

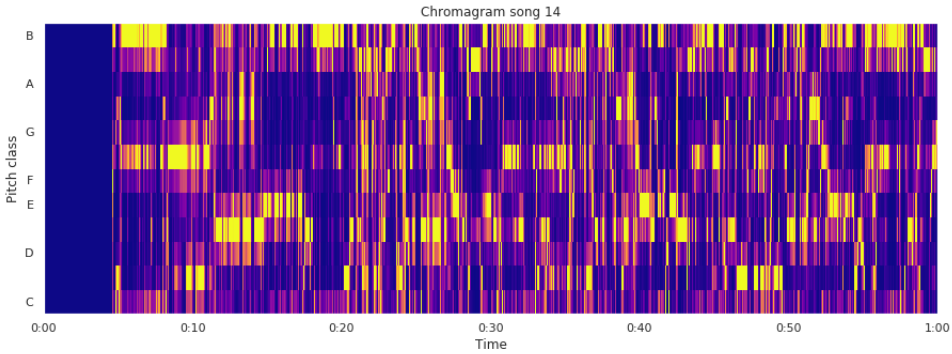

Figure C.2.9: Chromagram song 14.

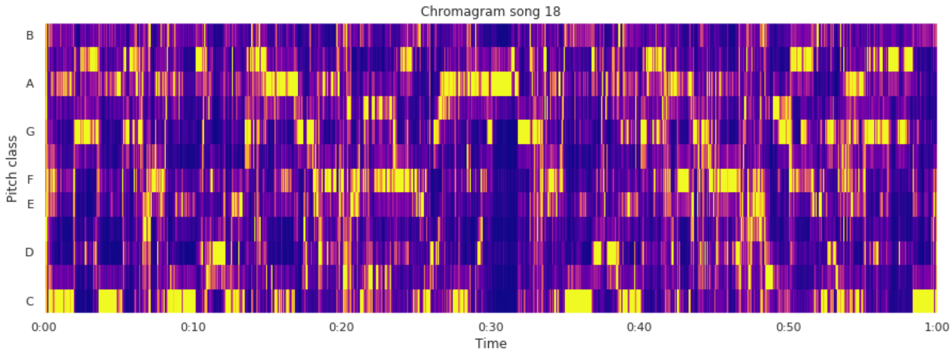

Figure C.2.10: Chromagram song 18.

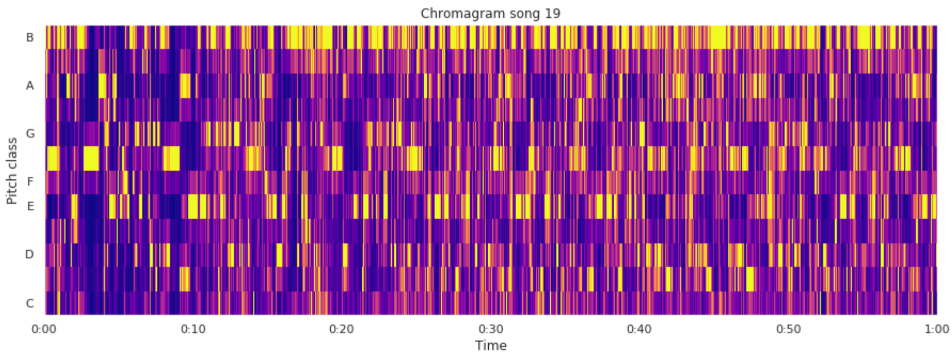

Figure C.2.11: Chromagram song 19.

C.3. Excited

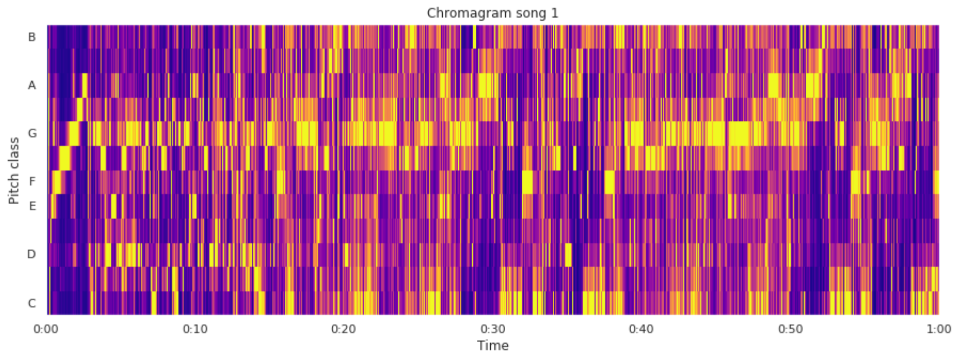

Figure C.3.1: Chromagram song 2.

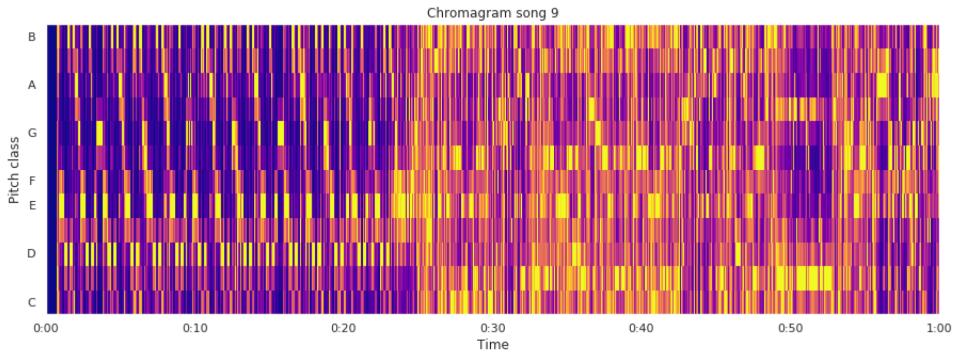

Figure C.3.2: Chromagram song 10.

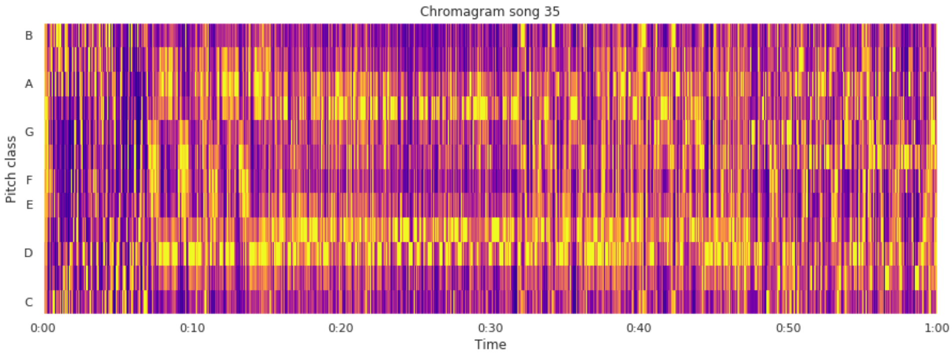

Figure C.3.3: Chromagram song 36.

C.4. Class neutral

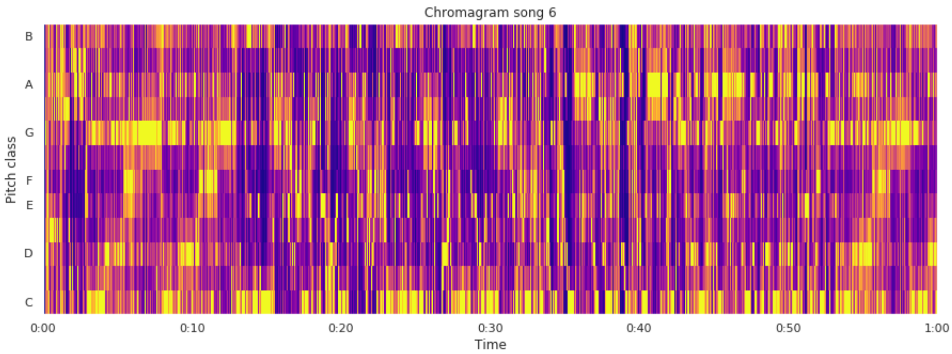

Figure C.4.1: Chromagram song 6.

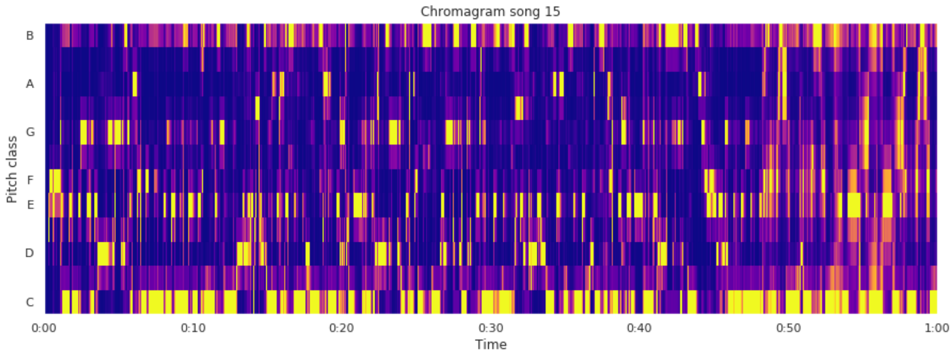

Figure C.4.2: Chromagram song 15.

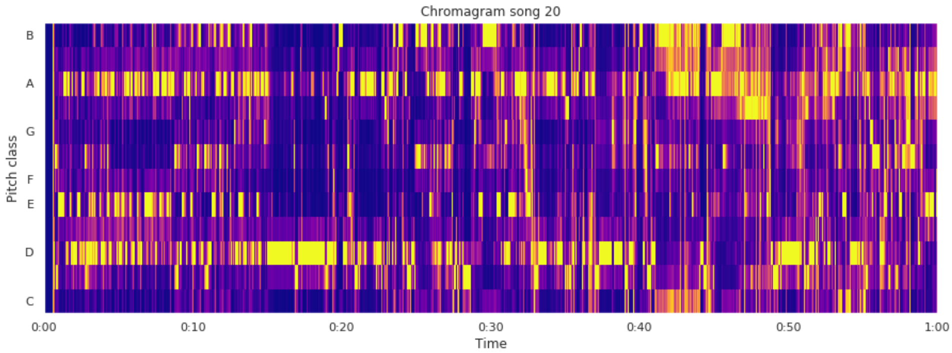

Figure C.4.3: Chromagram song 20.

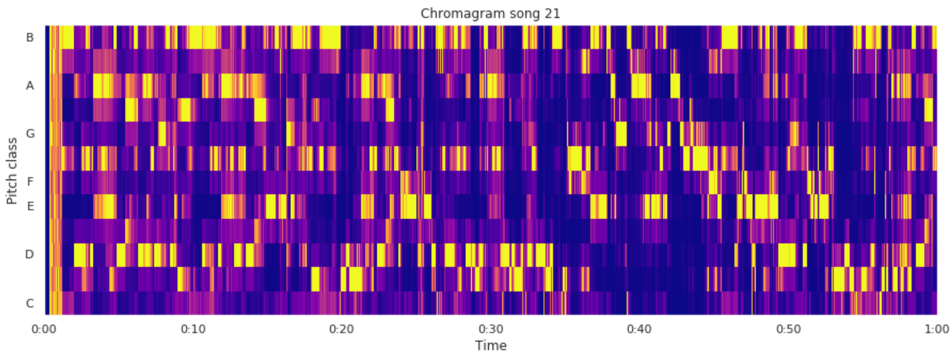

Figure C.4.4: Chromagram song 21.

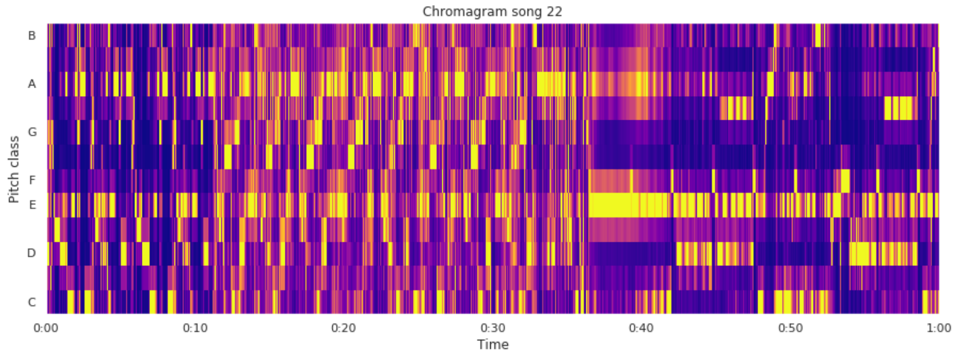

Figure C.4.5: Chromagram song 22.

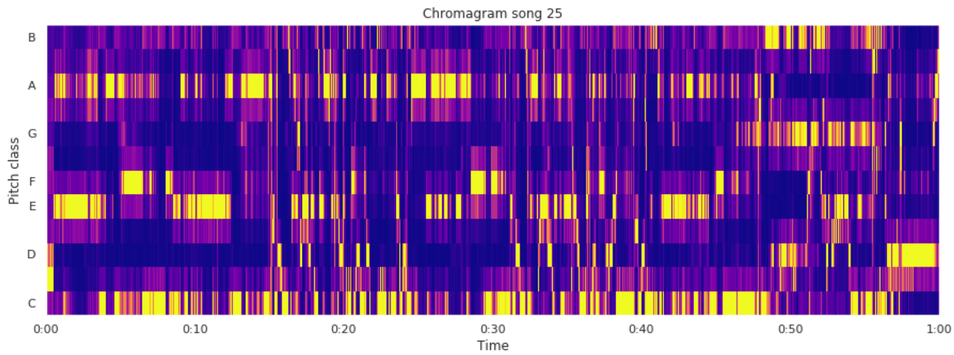

Figure C.4.6: Chromagram song 25.

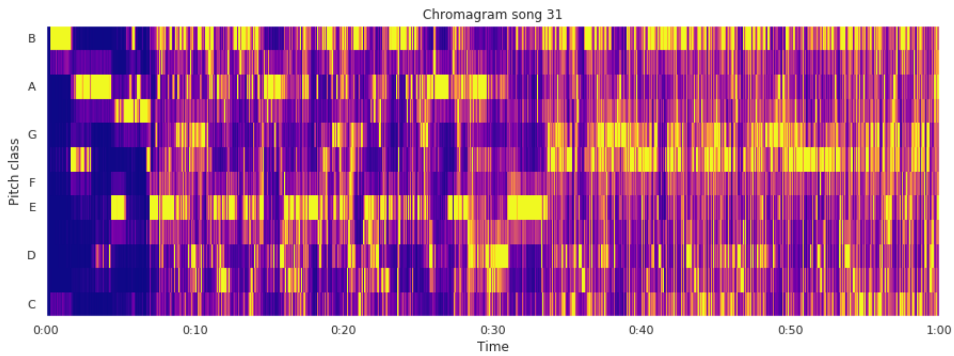

Figure C.4.7: Chromagram song 31.

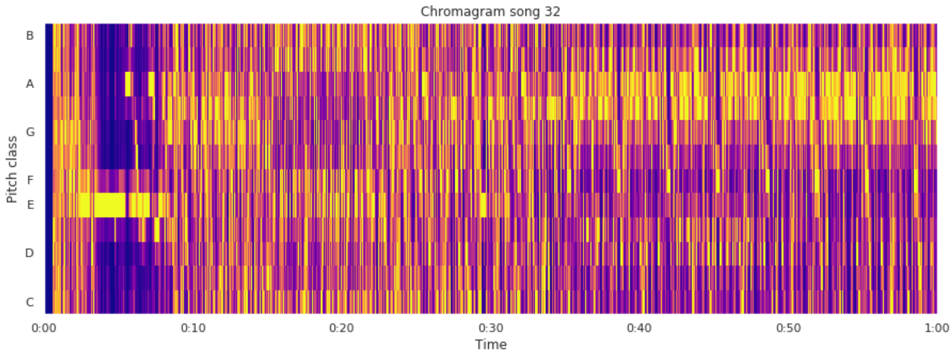

Figure C.4.8: Chromagram song 32.

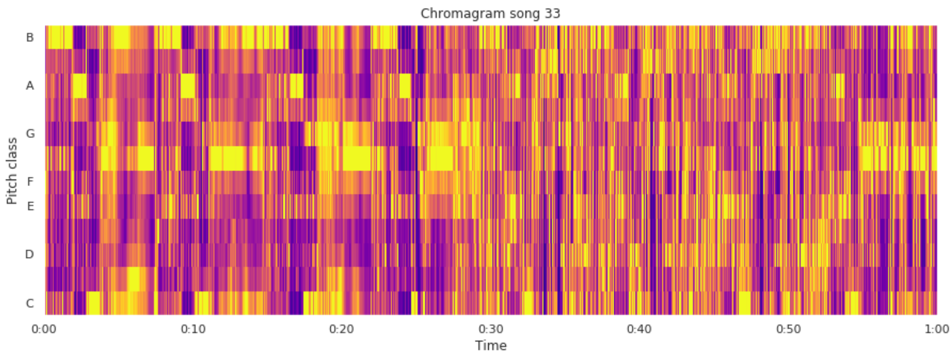

Figure C.4.9: Chromagram song 33.

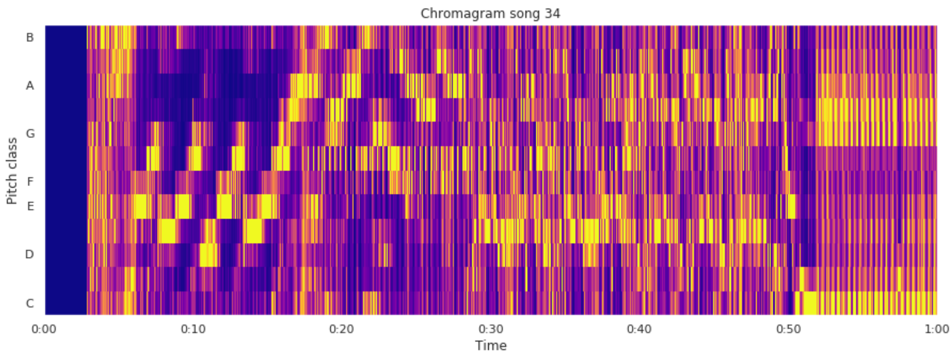

Figure C.4.10: Chromagram song 34.

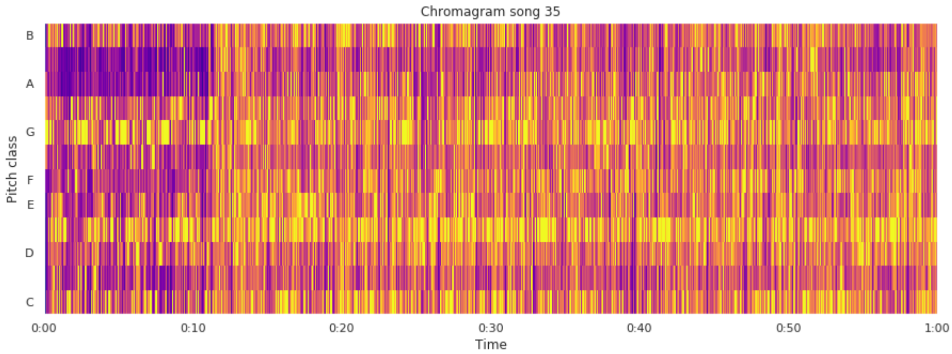

Figure C.4.11: Chromagram song 35.

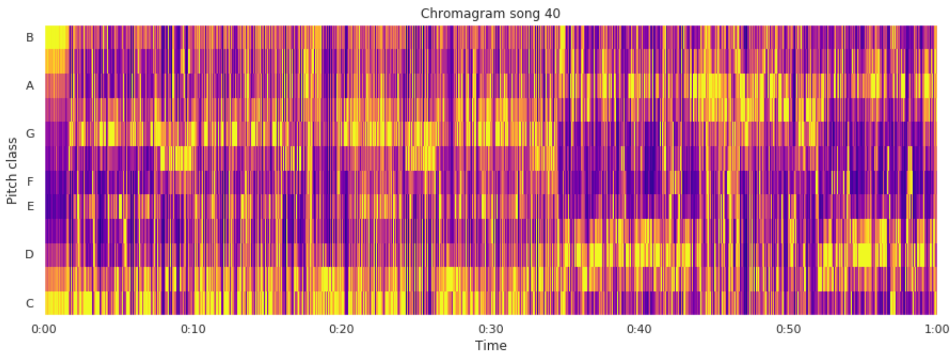

Figure C.4.12: Chromagram song 40.

C.5. Class calm

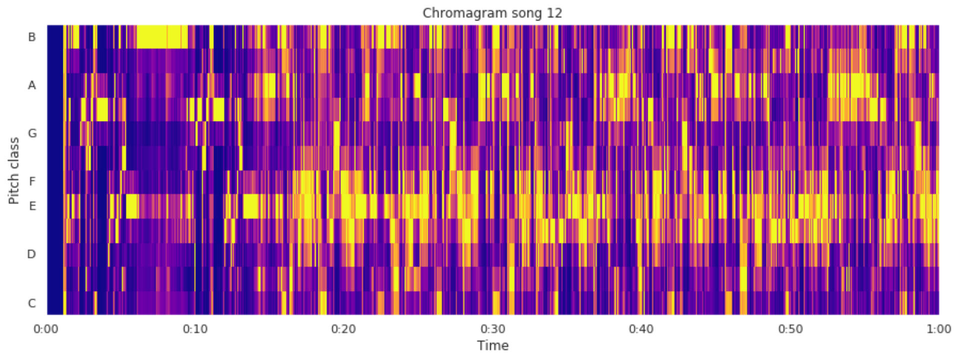

Figure C.5.1: Chromagram song 12.

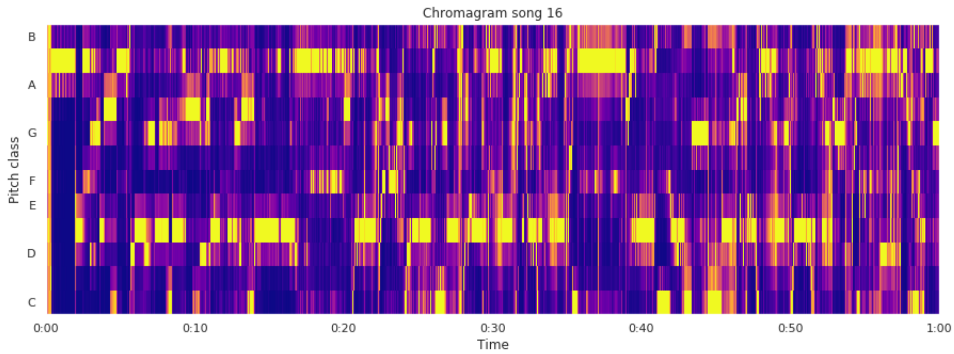

Figure C.5.2: Chromagram song 16.

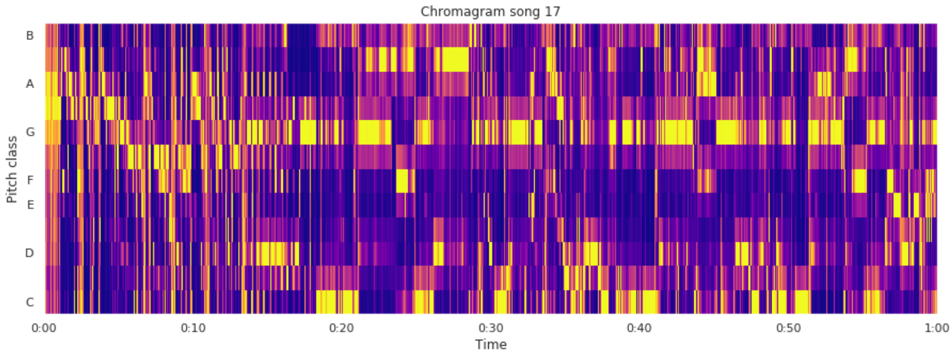

Figure C.5.3: Chromagram song 17.

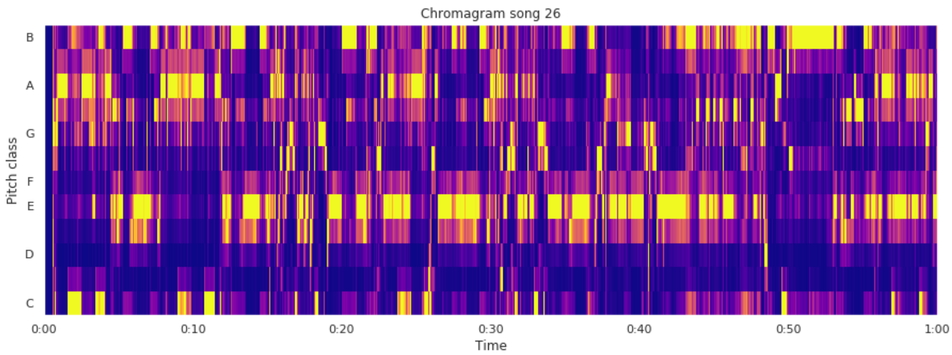

Figure C.5.4: Chromagram song 26.

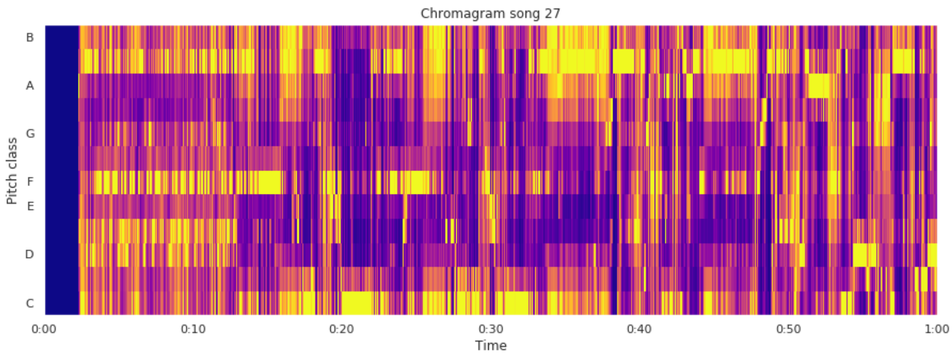

Figure C.5.5: Chromagram song 27.

C.6. Class miserable

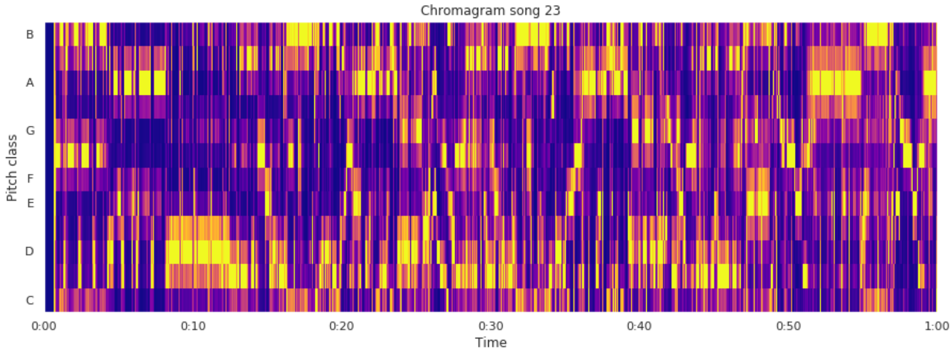

Figure C.6.1: Chromagram song 23.

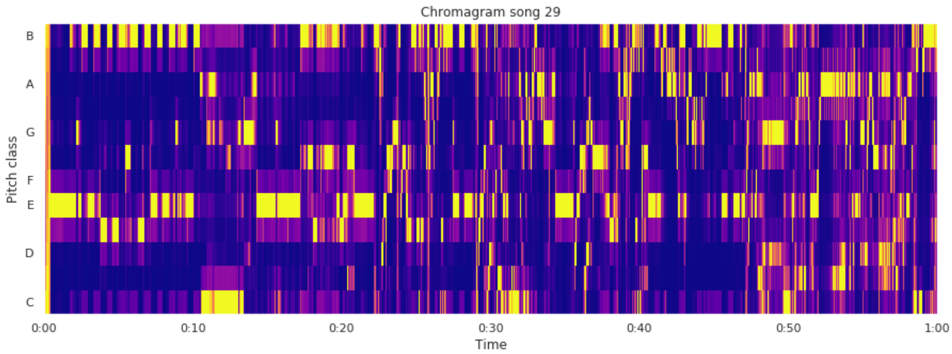

Figure C.6.2: Chromagram song 29.

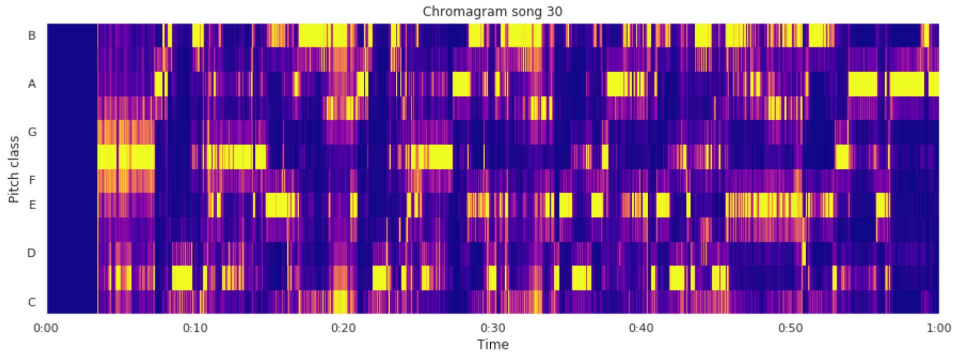

Figure C.6.3: Chromagram song 30.

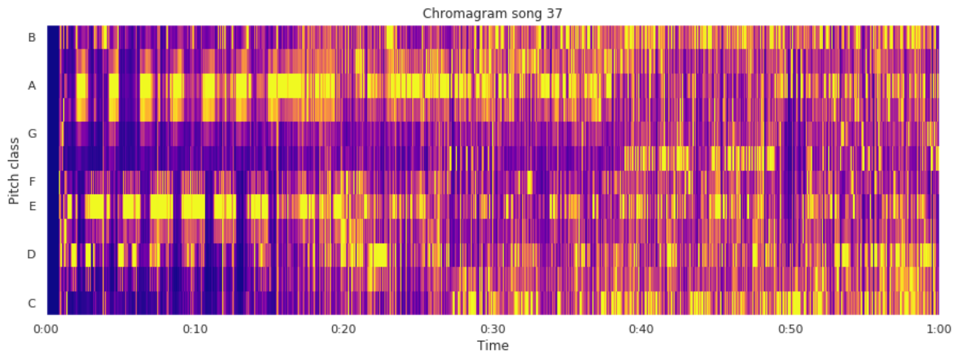

Figure C.6.4: Chromagram song 37.

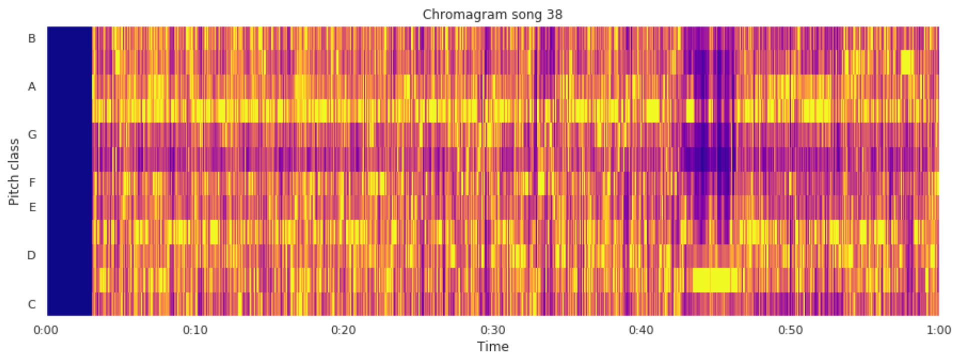

Figure C.6.5: Chromagram song 38.

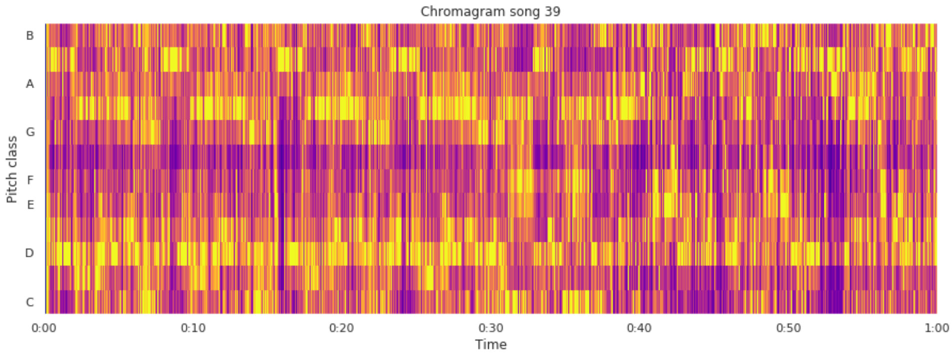

Figure C.6.6: Chromagram song 39.

C.7. Class depressed

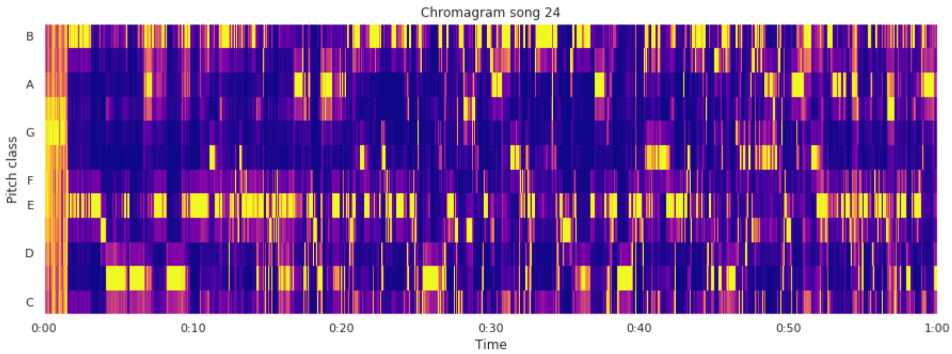

Figure C.7.1: Chromagram song 24.

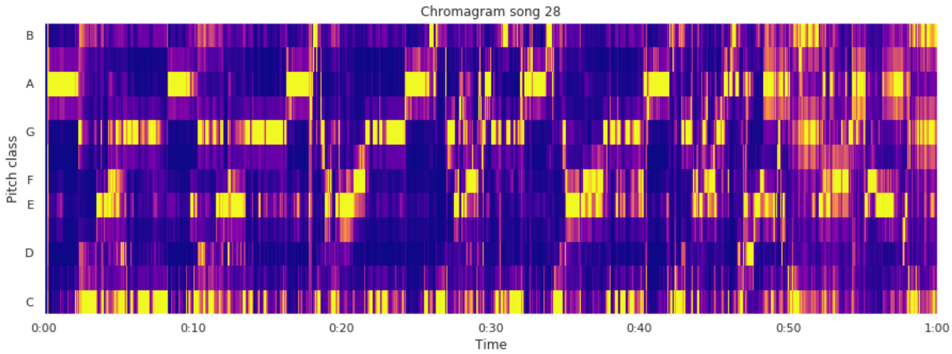

Figure C.7.2: Chromagram song 28.