Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Measuring the Evolution of Timbre in Billboard Hot 100

Degree's Thesis
Audiovisual Systems Engineering

Author: Aleu Pons Albà
Advisors: Jordi Pons Puig and Antonio Bonafonte Cávez

Universitat Politècnica de Catalunya (UPC)
2016 - 2017

Abstract

This project consists in analyzing the timbre blend of some representative most popular songs along last 60 years: Billboard Hot 100's first positions of all weeks of the analysed years. The study focus the attention in those years that made a revolutionary change in Western popular music, according to musicological bibliography.

After studying musical timbre and its state of the art in order to define the methodology, the data-set is defined, obtained and perceptually analysed and classified.

Then a two-parts system is developed. First part consist in timbre characterization using MFCCs (Mel-frequency cepstral coefficients), ZCC (zero-crossing count) and brightness musical descriptor. And second part is the automatic classification of all data-set using trained and validated SVM (support vector machine) models.

Finally, obtained results show an approximation of mainstream music timbre evolution. Despite of their questionable validity and the risks of falling into false attributions or generalizations, final conclusions point -with regard to Billboard Hot 100- to the non-influence of punk, an increase of high load of low frequencies and electronic sounds along time, and the influence of hip hop and trap.

Resum

Aquest projecte consisteix en analitzar la mescla tímbrica d'algunes de les cançons més populars al llarg dels últims 60 anys: les primeres posicions de la *Billboard Hot 100* de totes les setmanes dels anys analitzats. L'estudi centra l'atenció en aquells anys que van provocar un canvi revolucionari en la música popular occidental, segons la bibliografia musicològica.

Després d'estudiar el timbre musical i el seu estat de la qüestió per definir la metodologia, es defineix i s'obté el conjunt de dades i s'analitza i es classifica perceptualment.

Després es desenvolupa un sistema de dues parts. La primera part consisteix en la caracterització del timbre mitjançant MFCCs (coeficients cepstrals en les freqüències de Mel), ZCC (número d'encreuaments per zero) i el descriptor musical de brillantor. I la segona part és la classificació automàtica de tot el conjunt de dades utilitzant models SVM (màquina de vector de suport) entrenats i validats.

Finalment, els resultats obtinguts mostren una aproximació de l'evolució del timbre de la música *mainstream* (del corrent principal). Tot i la seva qüestionable validesa i els riscos de caure en atribucions errònies o generalitzacions, les conclusions finals apunten -respecte a la *Billboard Hot 100*- a la no-influència del punk, un augment de l'alta càrrega de baixes freqüències i sons electrònics al llarg del temps i la influència del *hip hop* i el *trap*.

Resumen

Este proyecto consiste en analizar la mezcla de timbre de algunas de las canciones más populares a lo largo de los últimos 60 años: las primeras posiciones de la *Billboard Hot 100* de todas las semanas de los años analizados. El estudio centra la atención en aquellos años que hicieron un cambio revolucionario en la música popular occidental, según la bibliografía musicológica.

Después de estudiar el timbre musical y su estado de la cuestión para definir la metodología, el conjunto de datos se define, se obtiene y perceptivamente se analiza y se clasifica.

Luego se desarrolla un sistema de dos partes. La primera parte consiste en la caracterización del timbre mediante MFCCs (coeficientes cepstrales en las frecuencias de Mel), ZCC (conteo de cruce a cero) y el descriptor musical de brillo. Y la segunda parte es la clasificación automática de todos los conjuntos de datos utilizando modelos SVM (máquinas de soporte vectorial) formados y validados.

Finalmente, los resultados obtenidos muestran una aproximación de la evolución del timbre de la música *mainstream* (de la corriente principal). A pesar de su cuestionable validez y los riesgos de caer en falsas atribuciones o generalizaciones, las conclusiones finales apuntan -con respecto a la *Billboard Hot 100*- a la no-influencia del punk, un aumento de alta carga de bajas frecuencias y sonidos electrónicos a lo largo del tiempo y la influencia del *hip hop* y el *trap*.

To music: amb tu va començar el canvi i t'ho agraeixo analitzant-te, respectant-te i reivindicant-te.

To those who despair: recordeu-me que la meva desesperació és afortunada, i que la canalitzi amb esperances col·lectives.

To all my friends: gràcies per les nocturnes hores de debat, el pensament crític i la reflexió constant. L'entorn que hem conformat ens modela i ens guia la conducta allà on anem.

To *Paraula 303*: ets l'inici d'un camí, les ganes i la il·lusió, la demostració que les alternatives no són només utopies.

Acknowledgements

First of all I would like to thank Jordi Pons his disinterested help, advices and knowledge he has given me during all processes of this thesis.

I would like to also thank my UPC advisor, Antonio Bonafonte, his help and flexibility since the first day when I purposed him my initial project.

And I do not want to neglect the initial point of view of Jordi Roquer. His brief contribution as musicologist was very helpful in order to focus and orient the study.

Finally, I would like to thank *MathWorks* community its *Matlab* software which has been a great and useful tool for the project. I specially appreciate the authors of *MIRtoolbox* and *RASTA-PLP* the open-source orientation and the potential of this toolbox.

Revision history and approval record

Revision	Date	Purpose
0	02/05/2017	Document creation
1	03/05/2017	Document revision
2	15/06/2017	Document revision
3	26/06/2017	Document approval

DOCUMENT DISTRIBUTION LIST

Name	e-mail
Aleu Pons Albà	aleu.pons@alu-etsetb.upc.edu
Jordi Pons Puig	idrojsnop@gmail.com
Antonio Bonafonte Cávez	antonio.bonafonte@upc.edu

Written by:		Reviewed by:		Reviewed and approved by:	
Date	02/05/2017	**Date**	03/05/2017	**Date**	26/06/2017
Name	Aleu Pons Albà	**Name**	Jordi Pons Puig	**Name**	Antonio Bonafonte Cávez
Position	Project Author	**Position**	Project Supervisor	**Position**	Project Supervisor

Contents

1 **Introduction** 13

1.1 Statement of purpose . . . 13

1.2 Requirements and specifications . . . 13

1.3 Methods and procedures . . . 14

1.4 Work plan . . . 15

1.4.1 Work packages . . . 15

1.4.2 Gantt diagram . . . 15

1.5 Incidents and modifications . . . 16

2 **State of the art** 17

2.1 Musical timbre . . . 17

2.2 Timbre characterization and similarity . . . 18

2.2.1 Timbre study in music information retrieval (MIR) . . . 18

2.2.2 Main conclusions about timbre research . . . 18

2.3 Audio signals representation . . . 20

2.3.1 Mel-frequency cepstral coefficients . . . 20

2.3.2 Zero-crossing count . . . 21

2.3.3 Spectral centroid frequency . . . 22

2.4 Classification model . . . 22

2.4.1 Support vector machines . . . 22

2.4.2 Overfitting and underfitting . . . 23

2.5 The evolution of Western popular music . . . 24

2.5.1 The 1950s . . . 24

2.5.2 The 1960s . . . 25

2.5.3 The 1970s . . . 25

2.5.4 The 1980s 26

2.5.5 The 1990s 27

2.5.6 The 2000s 27

2.5.7 The 2010s 27

2.5.8 Main revolutions of Western popular music since 1950s 28

3 Methodology **29**

3.1 Data-set 29

3.2 Perceptual analysis 30

3.2.1 1962-1966 - British pop emergence 30

3.2.2 1975-1979 - Disco and punk popularity 30

3.2.3 1989-1993 - Hip hop's mainstream emergence 31

3.2.4 1997-2001 - Boy bands, girl bands and pop starlets emergence 31

3.2.5 2012-2016 - Second wave of trap 31

3.3 Timbre characterization 31

3.4 Timbre evolution 33

3.4.1 Training 33

3.4.2 Test/validation 34

3.4.3 Timbre evolution prediction 34

4 Results **35**

4.1 Timbre characterization 35

4.2 Evaluation of the timbre characteristics classifiers 35

4.2.1 Vocal accompaniment (choruses) characteristic 36

4.2.2 Backing keyboard instruments characteristic 36

4.2.3 Orchestral instruments characteristic 37

4.2.4 Distorted power chords characteristic 37

4.2.5 Drum machine characteristic 37

4.2.6 High load of low frequencies characteristic 38
4.2.7 Electronic backing characteristic 38
4.2.8 Late 1990s vocal accompaniment (choruses) characteristic 39
4.2.9 Sub-bass kick drums characteristic 39
4.2.10 Twitchy hi-hats characteristic 39
4.2.11 Best configurations for the music characteristics 40
4.3 Classification and timbre evolution 40
4.4 Discussion of the results 43

5 Budget **45**

6 Conclusions **46**

A Additional information about the work plan **48**
A.1 Work Packages and tasks 48
A.2 Milestones 51

B Additional information about state of the art **52**
B.1 Timbre study in music information retrieval (MIR) 52
B.1.1 MIR, music evolution and musical descriptors efficiency 53
B.2 Timbre study in music psychology 55

C Constructed data-set and its perceptual classification **57**

D Timbre characterization results: figures of two examples **70**

E Evaluation measures used **73**

List of Figures

2.1 MFCC features computation 21

3.1 Scheme of the developed methodology 29

3.2 BOF construction from computed MFCCs. Complete BOF also contains the features obtained from deltas, ZCC and brightness. The figure shows the procedure for one song file. 32

3.3 Features vector of each song 32

3.4 Scheme of training data 33

3.5 Scheme of SVMs 10-fold cross-validation process. It has to be iterated 10 times for X=1:10 34

4.1 Vocal accompaniment (choruses) evolution in Billboard Hot 100 41

4.2 Backing keyboard instruments, orchestral instruments and distorted power chords evolution in Billboard Hot 100 41

4.3 Drum machine and high load of low frequencies evolution in Billboard Hot 100 42

4.4 Electronic backing and late 1990s vocal accompainment (choruses) evolution in Billboard Hot 100 42

4.5 Sub-bass kick drums and twitchy hi-hats evolution in Billboard Hot 100 . . . 43

D.1 Comparison of MFCCs for two different songs. The first 14 MFCCs are computed for frames of 50ms and half overlapping. 70

D.2 Comparison of ZCC values along time for two different songs. They are computed for frames of 50ms and half overlapping. 71

D.3 Comparison of brightness evolution for two different songs. It is computed for frames of 50ms and half overlapping. 72

List of Tables

2.1 Differences between MIR and music psychology 19

4.1 Training and test measures of vocal accompaniment characteristic for different configurations of 1962-1966 data-set 36

4.2 Training and test measures of backing keyboard characteristic for different configurations of 1975-1979 data-set 36

4.3 Training and test measures of orchestral instruments characteristic for different configurations of 1975-1979 data-set 37

4.4 Training and test measures of distorted characteristic for different configurations of 1975-1979 data-set 37

4.5 Training and test measures of drum machine characteristic for different configurations of 1989-1993 data-set 38

4.6 Training and test measures of high load of low frequencies characteristic for different configurations of 1989-1993 data-set 38

4.7 Training and test measures of electronic backing characteristic for different configurations of 1997-2001 data-set 38

4.8 Training and test measures of late 1990s vocal accompaniment characteristic for different configurations of 1997-2001 data-set 39

4.9 Training and test measures of sub-bass kick drums characteristic for different configurations of 2012-2016 data-set 39

4.10 Training and test measures of twitchy hi-hats characteristic for different configurations of 2012-2016 data-set 40

4.11 Best configurations measures for the music characteristics 40

5.1 Estimated budget of the project 45

A.1 Work Package 1 48

A.2 Work Package 2 49

A.3 Work Package 3 49

A.4 Work Package 4 50

A.5 Work Package 5 50

A.6 Milestones 51

C.1 1962-1966 data-set and its perceptual classification 59

C.2 1975-1979 data-set and its perceptual classification 63

C.3 1989-1993 data-set and its perceptual classification 65

C.4 1997-2001 data-set and its perceptual classification 67

C.5 2012-2016 data-set and its perceptual classification 69

Chapter 1

Introduction

In the following chapter the statement of purpose of the project is presented. Then, its requirements and specifications are stated, as well as the methods used for the development of this work. Finally, the work planification with its tasks, incidents occurred during the realization of the project and deviations from the initial plan are explained.

1.1 Statement of purpose

The analysis of humanity cultural evolution is a task generally conducted by different scientific and humanistic areas such as sociology, anthropology or philosophy. But nowadays, with the emergence and progress of experimental science and technology, it is also possible to study this evolution from other points of view. Above all we can see it in cultural and artistic expressions more likely to be analysed with technology, such as music, cinema or photography. Therefore we should not be surprised if the study of culture, in many ways, is also being study subject for engineers, psychologists, biologists, etc. Some of these studies, following rigorous and scientific methods, and starting from a number of assumptions and/or thesis, try to look for patterns, descriptions, correlations and conclusions to characterize mathematically these cultural expressions and their evolution over time.

In this sense, the presented work is born from various hypothesis and objections which raise the issue -from a cultural point of view- of an actual mainstream music stagnation and homogenization. In this project mainstream music is considered as those music that usually appears on the radio and is well known to the general public. But this work also starts from the belief that there are many important and constant influences from Western popular music in mainstream music. And it is assumed that actually timbre blend is the most relevant aspect of current mainstream music.

The main objective of the project is to analyse the evolution of timbre in mainstream music in last 60 years and to relation it (or no) with the influence of most significant rises of Western popular music genres. With obtained results, it will be possible to find the evolution of influences, differences, similarities, homogenizations, etc., as regards the musical aspect of timbre, between each Western popular music revolution and mainstream music.

The main contribution of this study to the music technology field is a characterization of the evolution of music timbre focused in mainstream music.

1.2 Requirements and specifications

The main requirement of the project is to **analyse the evolution of timbre in Billboard Hot 100 music in last 60 years, attending to that historical moments which**

signified rises of Western popular music genres, and using timbre descriptors to characterize and classify music.

The specifications of this study are:

- A music data-set (set of songs) has to be designed and obtained in order to represent Billboard Hot 100 music in significant moments of popular music genres rises.
- Several timbre characterizations (binary class labels) have to be defined a priori.
- A perceptual analysis of data-set has to be done in order to classify the songs and obtain the ground truth.
- The musical descriptors of the implemented system have to extract timbre features of data-set.
- The system has to be able to classify training and test data with 80% of accuracy at least.
- The system has to classify all data-set in all class labels.

1.3 Methods and procedures

The presented study aims to characterize the evolution of timbre in Billboard Hot 100, attending to the influences of Western popular music and the hypothesis about an actual mainstream music stagnation and homogenization. To study the issue it is necessary to trace the evolution of Western popular music in last decades. Thus it will be possible to compare Billboard Hot 100 music productions (representative of US mainstream music chart) of different periods to observe timbral changes before and after a determined popular musical revolution.

The song data-set used is composed by the Billboard Hot 100's first positions of all weeks of the analysed years. And these years are those that made a revolutionary change in Western popular music, according to musicological studies.

The procedure used consists in define a priori several timbre characterizations (binary class labels) according to musicological information. Then the data-set is classified perceptually into these classes in order to obtain the ground truth. Finally, using Mel-frequency cepstral coefficients (MFCCs), zero-crossing count (ZCC) and brightness, the timbre is characterized. Then, a support vector machines (SVMs) system is developed, trained and tested with this features in order to classify all data-set into class labels.

The training, validation and classification systems have been developed using MATLAB software (R2012a - 7.14.0.739 version). Additionally, several audio descriptors of MIRTOOLBOX (1.6.1 version), a MATLAB toolbox, have been used [29].

This project has been carried out independently and autonomously by the author with the help of Jordi Pons from Music Technology Group (MTG) of the Universitat Pompeu Fabra (UPF), and supervised by Antonio Bonafonte from Signal Theory and Communications Department (TSC) of the Universitat Politècnica de Catalunya (UPC). The work starts from scratch but it is influenced by other previous investigations about musical evolution in terms of audiovisual engineering. It has been designed in late 2016 and early 2017 and has been realized during the 2017 Spring semester.

1.4 Work plan

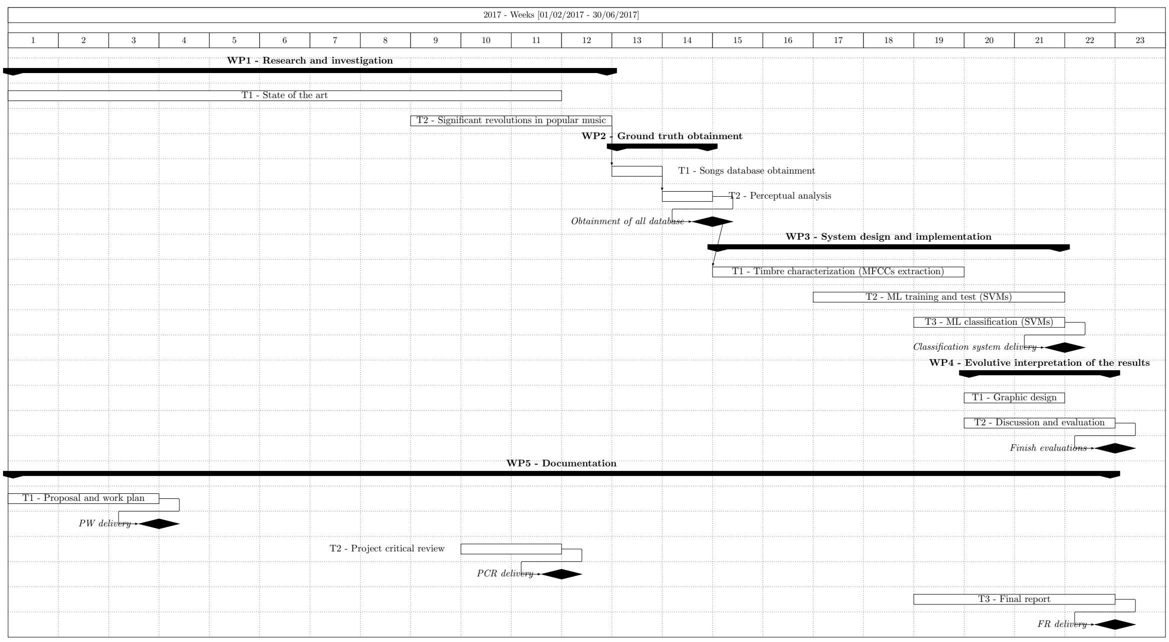
Initially the project was planned into four parts, but then another part was added, modifying the initial plan. The planning in this section corresponds to the latest one. The reasons for this change will be explained in section 1.5. A more detailed explanation of the work plan can be found in appendix A.

1.4.1 Work packages

- **WP1** - Research and investigation
- **WP2** - Ground truth obtainment
- **WP3** - System design and implementation
- **WP4** - Evolutive interpretation of the result
- **WP5** - Documentation

These Work Packages (WPs) are detailed in the Gantt diagram of section 1.4.2. Each WP tasks are represented in the diagram.

1.4.2 Gantt diagram



1.5 Incidents and modifications

The main incidence of the project has been the delay of one month approximately in first WP: the investigation and research part of the study. There have also been many problems and reassessment in epistemological terms. This means that all methodology -selection of music database, feature extraction and algorithms or systems used- has to have a meaningful goal for the investigation of the field in question, the music evolution in this case. So the project had to have clear the why and the sense when it implements the system and extracts music features.

It is also important to take into account the lack or inexperience of work planing that has delayed the tasks too. In this sense the obtainment of music database took much more time than expected.

The work plan was been modified according to the delay incidences and the new reorientations of the project. The obtainment of the specific database -a difficult, heavy and long task- and its perceptual analysis produced a new WP respect to the original plan.

Apart from that, in technical aspects of the work, the final decision was to focus on timbre feature extraction by using MFCCs, ZCC and brightness, and to classify music database with machine learning approach using SVMs. So the comparison between different music periods will be the comparison, the evolution, of its timbre features automatically classified and based on a ground truth perceptually made.

Chapter 2

State of the art

The study of the musical timbre, as well as many other related fields in music analysis, has been a reason of research for various audiovisual engineers in previous works. Music information retrieval (MIR) is the interdisciplinary science for excellence which tasks are very closed to the study of music -and timbre analysis in particular- from an engineering point of view. But there are other disciplines that involve similar studies of the same issues from another point of view, such as musicology, music psychology, signal processing, etc. So in order to know previous investigations of music timbre it is necessary to approach many different bibliography. In this way this will help knowing the engineering systems previously used to analyse, characterize and compare music timbre, and the main conclusions about timbre study. Therefore the objective of this state of the art is to obtain an optimal way to characterize and classify music timbre.

In this chapter a research of the evolution of Western popular music in last decades is also made. In this way, it will be possible to focus the attention into certain historical moments which produced notable changes or revolutions in mainstream trends of popular music. After that, it will be possible to analyse and compare timbre of mainstream songs in these years.

2.1 Musical timbre

First of all musical timbre has to be defined. According to Stephen McAdams definition [35], timbre is a set of auditory attributes that endows musical sounds with their particular 'colour', 'shape' or 'texture' (which may co-vary with pitch, loudness, and duration) and that enables listeners to identify sound sources. In [52] David L. Wessel also attribute quality of sounds to musical timbre.

In [8], Sophie Donnadieu wrote that timbre is a poorly understood auditory attribute, a strange and multiple attribute of sound defined by what is not: it is neither pitch, nor loudness, nor duration. She concludes that timbre is the perceptual music attribute by which we can distinguish the instruments of the orchestra even if they play the same note (same pitch) with the same dynamics (same loudness).

But there are also more definitions and different approximations to the term. As we can see in [40], the term *timbre* encompasses a set of auditory attributes of sound events in addition to pitch, loudness, duration, and spatial position. While psychoacoustic research has modeled timbre as a multidimensional phenomenon and represents its perceptual structure in terms of *timbre spaces*, in MIR field, perceptually relevant timbre parameters are needed as indices for content-based search of targeted timbres in very large sound databases, as well as for automatic categorization, recognition and identification schemes for musical instrument and environmental sounds.

2.2 Timbre characterization and similarity

Once exposed the principal definitions of musical timbre, it is necessary to focus attention on methods that have been used before to characterize and to compare musical timbre. These methods can be divided principally in two groups, according to the discipline which develop them: MIR or music psychology. In this section only a brief introduction to timbre study in MIR, and main conclusions about timbre research will be presented. See appendix B for complete information.

2.2.1 Timbre study in music information retrieval (MIR)

Timbre study in MIR is related to audio-based tasks, sound source recognition and instrument classification. Normally, they are divided into a two-parts algorithm:

- Audio signals representation for which audio descriptors are chosen or are newly created, generally obtained by mapping the signal's short-term Fourier transform (STFT) into a lower-dimensional domain that shows more clearly the relevant signal characteristics. Sometimes it is followed by a feature selection that removes the least important descriptors from the model.
- Selection of the classification model, such as k-nearest neighbours (k-NN), Gaussian mixture models (GMM), hidden Markov models (HMM), support vector machines (SVM) or neural networks (NN). Then, the model is trained and evaluated, usually by employing cross-validation with regards to a *ground-truth* defined in advance on annotated sets of audio data.

Several previous MIR studies have developed different systems using algorithms schemes like the presented before, in order to classify or recognize instruments by characterizing timbre directly or indirectly. Some of them are presented at appendix B.1.

2.2.2 Main conclusions about timbre research

Below several conclusions about music timbre investigations related to articles and bibliography researched are exposed in order to approach to principal and general conclusions founded.

First of all, the differences between methods, techniques, tasks and objectives pursued by principal timbre research fields have to be exposed. In this way it will be possible to approach to the most useful investigations' conclusions for the purpose of this project. Kai Siedenburg et al. analyses the different usage of audio descriptors for timbre research in MIR and music psychology [47]. Principal differences between these disciplines are exposed in table 2.1

The conclusion of this study outlines differences and epistemological foundations of the two fields: they are studying different facets of timbre-related tasks in most of the cases. MIR most often considers classification while music psychology has predominantly dealt

	MIR	Music psychology
Tasks	Automatic instrument classification	Study timbre perception by relying on dissimilarity ratings
Audio descriptors	Uses multitude of audio descriptors (>20)	Uses only highly constrained set of descriptors (<5)
Obtained knowledge	No clear epistemological status	Describes the physical correlates of timbre perception

Table 2.1: Differences between MIR and music psychology

with timbre dissimilarity perception. The authors argues that as cognitive phenomena, it is by no means clear that classification and similarity assessment rely on the same compilation of perceptual features when dealing with high-dimensional perceptual objects such as timbre.

In the same sense, a set of problems that hinders the academic disciplines of MIR and music cognition from productive collaboration were identified [1]. Authors cover the usage of descriptors and algorithms and methods for scientific validation, as well as the status of limit cases in both fields. Their most fundamental claim is that parts of MIR are methodologically misguided because they do not respect the ways in which auditory information processing takes place in humans, but rather employ heuristics that have proven to be successful when evaluated algorithmically over large corpora of music. In the case of successful evaluation, these heuristics are then falsely interpreted as evidence of mechanisms of human auditory information processing.

Bob Sturm also argued that relying on classification accuracy alone can be a misleading criterion in tasks such as musical genre classification, sometimes giving rise to operationalizations of musical genre that are implausible to any human listener [50, 49].

These findings imply that in many situations there may indeed be a much smaller number of substantially independent variables present than raw numbers of descriptors suggest.

Secondly, guidelines provided by other authors can be followed in order to take into account those essential timbre characteristics which a good system have to characterize. But also it is important to take into account the attributes which timbre has to be invariant, according to [42] and timbre definition exposed at the beginning of this section:

- Pitch invariance
- Loudness invariance
- Duration invariance
- Spatial position invariance

Finally, in conclusion, it is important to focus attention in few of the audio descriptors and features which most of the previous works conclude that are important in terms of characterizing musical timbre. So according that, three music descriptors were selected. Then, they were used to classify binarily all data-set into 10 representative music characteristics designed and validated perceptually (see section 3.2). Next sections expose these selected features.

2.3 Audio signals representation

According to the two-parts algorithm exposed at section 2.2.1, the first part of the system is the audio signals representation with the selected audio descriptors or features. The decision was to use MFCCs coefficients to describe timbre of the data-set, along with zero-crossing count (also: zero-crossing rate) and brightness (understood as spectral centroid frequency or center of gravity of the spectrum).

2.3.1 Mel-frequency cepstral coefficients

This section has been developed after consulting [30], so much of the explanation set out below is derived from cited article.

Mel-frequency cepstral coefficients (MFCCs) have been the dominant features used for speech recognition for some time. Their success has been due to their ability to represent a description of the spectral envelope shape of the sound in a compact form, so they seem optimal for timbre characterization (i.e. [12, 24, 34]). Each step in the process of creating MFCC features is motivated by perceptual or computational considerations. As representation of the spectral envelope shape and its energy and frequency distributions have been recurrently and majority used and assumed as timbre representative features in previous studies (see section 2.2 and appendix B), MFCCs have been used in this work in order to characterize the blend timbre.

The first step is to divide the audio waveform (speech signal or music recording) into frames, usually by applying a windowing function at fixed intervals. The aim here is to model small sections of the signal that are statistically stationary. A cepstral feature vector for each frame is generated.

Next step consist in taking the Discrete Fourier Transform (DFT) of each frame (2.1, 2.2). Then, only the logarithm of the amplitude of the spectrum is retained (2.3) because the perceived loudness of a signal has been found to be approximately logarithmic. Phase information is discarded because perceptual studies have shown that the amplitude of the spectrum is much more important than the phase.

$$\mathcal{F}\{x\} = X; \quad x = x_0, x_1, ..., x_{N-1}; \quad X = X_0, X_1, ..., X_{N-1} \tag{2.1}$$

$$X_k = \sum_{n=0}^{N-1} x_n \;\; \mathrm{e}^{-j2\pi kn/N} = \sum_{n=0}^{N-1} x_n \;\; [\cos(2\pi kn/N) - j\sin(2\pi kn/N)] \tag{2.2}$$

$$S_k = \log(|X_k|) \tag{2.3}$$

After that, the spectrum is smoothed and meaningful frequencies are emphasized perceptually. This process is achieved by collecting the spectral components into a lower number of frequency bins using F *Mel* filters. These bins are not equally spaced in frequency due to the perceptually more importance of lower frequencies than higher frequencies. Therefore,

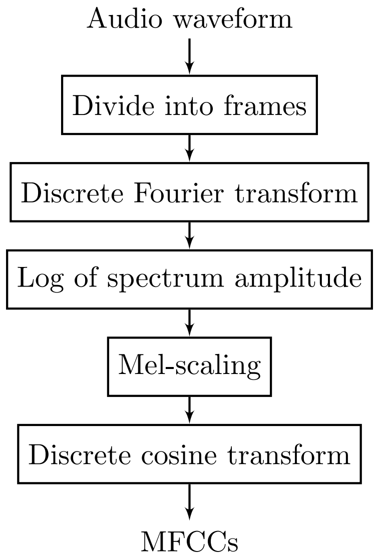


Figure 2.1: MFCC features computation

the bin spacing follows the so-called Mel frequency scale. The Mel scale is a perceptual scale of pitches judged by listeners to be equal in distance from one another, based on a mapping between actual frequency and perceived pitch as apparently the human auditory system does not perceive the pitch in a linear manner. The mapping is approximately linear below 1kHz and logarithmic above. A popular formula to convert f hertz into m mels is 2.4. Every Mel bin or band has and associated energy E_k, where $k = 0, ..., F-1$ and F is the number of filters.

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \tag{2.4}$$

The components of the Mel-spectral vectors calculated for each frame are highly correlated. In order to reduce the number of parameters in the system the last step of MFCC feature construction is to apply a transform to the Mel-spectral vectors which decorrelates their components. Theoretically, Karhunen-Loeve (KL) transform achieves this, but it is usually approximated by Discrete Cosine Transform (DCT) [38], obtaining L (13 or so) cepstral features for each frame (see 2.5).

$$C_k = \sum_{n=0}^{F-1} E_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right]; \quad 0 \le k < L \tag{2.5}$$

2.3.2 Zero-crossing count

The Zero-crossing count (ZCC or zero-crossing rate, ZCR) is a time-domain audio feature which has been used in speech recognition [21] and has been applied successfully to discern drum sounds [16]. It is calculated by simply counting the number of times consecutive samples in a frame are of opposite sign. ZCC is high for noisy signals and transient sounds at the onset of consonants and percussive events. ZCC is defined formally as 2.6 formula.

$$ZCC = \frac{1}{T-1}\sum_{t=1}^{T-1} 1_{R_{<0}}(s_t s_{t-1}) \tag{2.6}$$

2.3.3 Spectral centroid frequency

Spectral centroid frequency is the measure used to characterize the sound frequency spectrum. It indicates where the center of gravity of the spectrum is. Perceptually, it has a robust connection with the impression of *brightness* of a sound [19, 23]. It is calculated as the weighted mean of the frequencies present in the signal, determined using Fourier transform, with their magnitudes as the weights:

$$Centroid = \frac{\sum_{k=0}^{N-1} f_k |X_k|}{\sum_{k=0}^{N-1} |X_k|} \tag{2.7}$$

where X_k represents the weighted frequency value of bin or frequency band k and f_k represents the center frequency of that bin.

Another interpretation of sound brightness correlated to spectral centroid frequency calculates it as the amount of spectral energy corresponding to frequencies higher than a given cut-off threshold (the spectral centroid frequency or one value around) [29]. This brightness computation method is proportioned by MIRtoolbox and it is the selected to complement MFCCs and ZCC to describe timbre of the songs of the project (see chapter 3).

2.4 Classification model

The second part of the system (see section 2.2.1) is the classification model. In this case the selected one has been support vector machine.

2.4.1 Support vector machines

The following explanation is based in [15].

Support vector machines (SVMs) [6] were introduced by Vapnik and co-workers [3] and successively extended by a number of other researchers. SVMs are a relatively new type of machine learning algorithm, and their remarkably robust performance with respect to sparse and noisy data is making them the system of choice in a large number of applications.

When used for classification, they separate a given set of binary labeled training data with hyper-plane that is maximally distant from them. For cases in which no linear separation is possible, they can work in combination with the technique of *kernels*, that automatically applies a non-linear mapping to a feature space. The hyper-plan found by SVM in feature space corresponds to a non-linear decision boundary in the input space.

If jth input point $\mathbf{x}^j = (x_1^j, ..., x_n^j)$ is the realization of the random vector $\mathbf{X}^j$ labeled by the random variable $Y^j \in \{-1, +1\}$, and $\phi : I \subseteq \mathcal{R}^n \to F \subseteq \mathcal{R}^N$ is a mapping from the input space $I \subseteq \mathcal{R}^n$ to a feature space F, let assume sample S of m labeled data points: $S = \{(\mathbf{x}^1, y^1), ..., (\mathbf{x}^m, y^m)\}$. The SVM learning algorithm finds a hyper-plane $(\mathbf{w}, b)$ such that the quantity

$$\gamma = \min_i y^i \{\langle \mathbf{w}, \phi(\mathbf{x}^i) \rangle - b\} \tag{2.8}$$

is maximized, where $\langle , \rangle$ denotes an inner product, the vector $\mathbf{w}$ has the same dimensionality as F, $\|\mathbf{w}\|_2$ is held constant, b is a real number, and γ is called the *margin*. The quantity $\{\langle \mathbf{w}, \phi(\mathbf{x}^i) \rangle - b\}$ corresponds to the distance between the point $\mathbf{x}^i$ and the decision boundary. When multiplied by the label y^i, it gives a positive value for all correct classifications and a negative value for the incorrect ones. The minimum of this quantity over all the data is positive if the data is linearly separable, and is called the margin. Given a new data point $\mathbf{x}$ to classify, a label is assigned according to its relationship to the decision boundary, and the corresponding decision function is:

$$f(\mathbf{x}) = sign(\langle \mathbf{w}, \phi(\mathbf{x}) \rangle - b) \tag{2.9}$$

The decision function can equivalently be expressed (2.11) with maximal margin hyper-plane (2.10) formed by *support vectors* (α_i), m positive real numbers. These points lie closest to the separating hyper-plane. And the matrix $K_{ij} = \langle \phi(\mathbf{x}^i), \phi(\mathbf{x}^j) \rangle$ is called the *kernel matrix* and is particularly important in some extensions of the algorithm when the data are not linearly separable. To provide non-linear decision boundaries several kernel functions can be used: polynomial, Gaussian, radial basis function or multilayer perception. These functions will be discussed and tested in next chapters.

$$\mathbf{w} = \sum_{i=1}^{m} \alpha_i y_i \phi(\mathbf{x}^i) \tag{2.10}$$

$$f(\mathbf{x}) = sign\left(\sum_{i=1}^{m} \alpha_i y_i \langle \phi(\mathbf{x}^i), \phi(\mathbf{x}) \rangle - b \right) \tag{2.11}$$

Robust performance with sparse data and its capability to classify binary labeled data make SVMs as the optimal classification model for the system purposed in chapter 3.

2.4.2 Overfitting and underfitting

One recurrent problem with machine learning or statistics classification models is the *overfitting*. But *underfitting* also can take place.

The principle of parsimony (or Occam's Razor) calls for using models and procedures that contain all necessary features to obtain a model but nothing else. For example, if a regression model with two predictors is enough to explain one success, then no more than these two predictors should be used. Going further, if the relationship can be captured by a linear function in these two predictors, then using a quadratic function violates parsimony. Overfitting is the use of models or procedures that violate parsimony, that is, that include more terms than are necessary or use more complicated approaches than are necessary [20]. But using an excessive amount of data to train models also could produce overfitting because it could be redundant.

When a statistical model or machine learning algorithm cannot capture the underlying trend of the data then the underfitting is taking place. Underfitting occurs when the model or algorithm does not fit the data enough, in other words, it occurs if the model or algorithm shows low variance but high bias (to contrast the opposite, overfitting occurs when the model or algorithm shows high variance and low bias) [4]. It is often a result of an excessively simple model or a model obtained with insufficient number of observations (few data).

2.5 The evolution of Western popular music

This section has been developed after consulting [32], so much of the explanation set out below is derived from cited chapter.

Once researched on musical timbre studies in order to characterize the technical part of the work, it is necessary to search in musicological and sociological bibliography the evolution of Western popular music in last decades. In this way, it will be possible to focus attention into certain historical moments which produced notable changes or revolutions in mainstream trends of popular music.

The historical review begins in 1950s because of the television introduction into people's homes and its revolutionary change provoked in how music and culture began to be consumed by masses. Another important reason is the birth of different mainstream music record charts such as US Billboard Hot 100 or UK Singles chart.

2.5.1 The 1950s

In the early 1950s the introduction of television into people's homes spread rapidly because of cheaper mass-production costs, so it threatened the existence of the radio industry. But this fact promulgated its adaptation by focusing on music, joining forces with the recording industry to survive and becoming a promotional tool. Instead of live music, recorded music gained presence in radio stations to fill airtime, and in 1955 the Top 40 format was born. Playlists for radio stations were based on popularity (usually the Billboard Top 40 singles chart). The purpose of this study is to focus the timbre analysis in this kind of music, the mainstream music which usually fill record charts like Billboard Hot 100, Top 40 or UK Singles Chart.

Back to the 1950s, television commercialization was not the only revolution that took place. Artists such as B. B. King surged in popularity with their urban Chicago blues. This genre spread among white and black teenagers alike and it was called as rhythm and blues (or R&B). Their sexually suggestive lyrics and the electrified guitar and wailing harmonica sounds appealed to young listeners [32]. In those times, R&B records were classified as *race music* and their sales were segregated from the white music records tracked on the pop charts [51].

One relevant moment was in 1951, when a late-night R&B show called *The Moondog Rock & Roll House Party* began referring as rock and roll to the R&B music played. The name comes from a blues slang term for sex. The early rock and roll music obtained instant notoriety, gaining widespread support among teenage music fans and widespread dislike among the older generation [41].

2.5.2 The 1960s

Rock and roll was mainly an American export before 1964. But this fact changed suddenly with the arrival of British pop phenomenon: The Beatles. Combining elements of skiffle -a type of music played on rudimentary instruments, such as banjos, guitars, or home-made instruments-, doo-wop, and soul, The Beatles popularized a genre of music known as Merseybeat (or beat music), named after the River Mersey. Merseybeat groups usually had simple guitar-dominated line-ups, with vocal harmonies and catchy tunes [46]. The most common instrumentation of beat groups featured lead, rhythm and bass guitars plus drums, as popularized by the Beatles, the Searchers, and others [31]. They often sang both verses and choruses in close harmony, resembling doo-wop, with nonsense syllables in the backing vocals [39]. The success and popularity of The Beatles in the United States extended to other British bands like The Rolling Stones, and by the mid-1960s they were all making appearances on the U.S. charts. Urban rock sound of The Rolling Stones was very different from pop music and remained more true to the bluesy, R&B roots of rock and roll. The called British Invasion transformed rock and roll into two different directions of future performers of this genre: the melodic, poppy sounds of the Beatles, on one hand, and the gritty, high-volume power rock of the Stones on the other [32]. But the branching out of rock and roll continued in several other directions throughout the 1960s (surf music, soul, folk rock...).

2.5.3 The 1970s

Musically, the ideological shift produced after the end of Vietnam War -when selfish views took the place of concern with social and political activism- resulted in the creation of glam rock, an extravagant, self-indulgent form of rock that incorporated flamboyant costumes, heavy makeup, and elements of hard rock and pop. The disco genre also emerged in the 1970s: flamboyant as glam rock but rising out of a more electronic sound. Thanks to the success of 1977 film *Saturday Night Fever*, disco popularity spread across USA. Record companies produced melodies with commercial objectives, so disco records were created especially for discos, becoming huge hits on the dance floor. The music typically layered soaring, often-reverberated vocals, often doubled by horns, over a background *pad* of electric pianos and *chicken-scratch* rhythm guitars played on an electric guitar. Other backing keyboard instruments include the piano, electric organ, string synth, and electromechanical keyboards. The sound was enriched with solo lines and harmony parts played by a variety of orchestral instruments, such as harp, violin, viola, cello, trumpet, saxophone, trombone, clarinet, flugelhorn, French horn, tuba, English horn, oboe, flute, piccolo, timpani and synth strings, string section or a full string orchestra [53].

Also in the 1970s, a new genre was created reacting against the commercialism of disco and corporate rock. Punk rock was a minimalist, angry form of rock that returned to rock and roll basics: simple chord structures, catchy tunes, and politically motivated lyrics [32]. The lure of punk rock, emerged out of a small bar in New York City, was that anyone with basic musical skills could participate. But it never reached a huge commercial success in the United States, instead of this, it exploded in the United Kingdom, where high unemployment rates and class divisions had created angry, disenfranchised youths. Punk rock vocals sometimes sound nasal, and lyrics are often shouted instead of sung in a conventional sense, particularly in hardcore styles [55]. Shifts in pitch, volume, or intonational style are relatively infrequent, and complicated guitar solos are considered unnecessary, although basic guitar breaks are common [5]. Guitar parts tend to include highly distorted power chords or barre chords. The first studio album by the English punk rock band Sex Pistols -*Never Mind the Bollocks, Here's the Sex Pistols*-, published in 1977, could be a culminating moment of the genre. In the late 1970s, punk bands began to abandon their sound when the genre became assimilated into the rock mainstream.

2.5.4 The 1980s

In the 1980s many disenfranchised black American youths expressed their displeasure by turning to hip hop -a term for the urban culture that includes break dancing, graffiti art, and the musical techniques of rapping, sampling, and scratching records-. The beginning of hip hop genre popularity was in the late 1970s among black youths, when record spinners in the Bronx and Harlem started to play short fragments of songs rather than the entire track (known as sampling). Firstly, early hip hop artists sampled all types of genres, like funk, soul and jazz. Later, they added special effects to the samples and experimented with techniques such as rotating or scratching records back and forth to create a rhythmic pattern. Also the DJs added short raps to their songs in order to let audiences know who was playing the records. This trend would then be further elaborated until entire spoken verses were included to the music. Hip hop's early evolution into a form distinct from R&B, not coincidentally, occurred around the time that sampling technology and drum-machines became widely available to the general public at a cost that was affordable to the average consumer (not just professional studios). A second wave of rap artists brought inner-city rap to American youths by mixing it with hard guitar rock in the early 1980s. Another subgenre that emerged was gangsta rap, a controversial brand of hip hop which highlights violence and gang warfare.

The Public Enemy's album *Fear of a Black Planet* played a key role in hip hop's mainstream emergence in 1990, dubbed by Billboard editor Paul Grein as "*the year that rap exploded*" [22]. In a 1990 article on its commercial breakthrough, Janice C. Simpson of *Time* wrote that hip hop "*grown into the most exciting development in American pop music in more than a decade*". Simpson noted the impact of Tone Lōc's single *Wild Thing* being the best-selling single of 1989, and that at the time of her article, nearly a third of the songs on the Billboard Hot 100 were hip hop songs [48].

2.5.5 The 1990s

Due to the maintained popularity of hip hop and gangsta rap in the early 1990s, in 1991 and 1992 several hip hop songs reached the first position of Billboard Hot 100. But also an alternative rock came to the forefront in the 1990s with grunge. The grunge scene emerged in the mid-1980s in the Seattle area of Washington State. Grunge was inspired by hardcore punk and heavy metal and it was so-called because of its messy, sludgy, distorted guitar sound, the disheveled appearance of its pioneers, and the disaffected nature of the artists [32]. Nirvana -developing a sound that relied on dynamic contrasts, often between quiet verses and loud, heavy choruses- is an example of a very success grunge band: after signing to major label DGC Records, they found unexpected success with *Smells Like Teen Spirit* (1991). It has to be said that more recently, alternative rock has fragmented into even more specific subgenres.

Simultaneously punk rock in the United States underwent a resurgence in the early to mid-1990s. Punk rock at that time was not commercially viable, and no major record label signed a punk rock band until Green Day's breakthrough in 1994.

In late 1990s, pop music was defining the mainstream tastes of the audience. A large number of boy bands, girl bands, and pop starlets -such as Britney Spears, Jennifer Lopez, Christina Aguilera, Backstreet Boys or The Spice Girls- emerged between 1997 and 2000, sometimes evolving from gospel choir groups, but more often than not created by talent scouts. These groups and pop starlets were aggressively marketed to teen audiences.

2.5.6 The 2000s

The 2000s began with pop music staying strong, until achieving the mainstream success throughout the decade. Some examples are country artists like Carrie Underwood and Taylor Swift, which by the end of the decade transitioned to good faith pop stars. Although rock music maintained its popularity at the beginning of the decade, its presence in mainstream music waned with a few exceptions by the end of the 2000s. Instead, hip hop maintained its popularity with more commercial artists achieving enormous success.

But this decade also saw the rise of reggaeton music genre. Reggaeton is a musical genre originated in Puerto Rico during the late 1990s in an underground scene. It has explicit lyrics about drugs, violence, poverty, friendship, love and sex, and it is characterized by its dembow beat (this created by Jamaican dancehall producers during the late 1980s and early 1990s). Dembow consists of a kick drum, kickdown drum, palito, snare drum, timbale, timbale roll and (sometimes) a high-hat cymbal. In 2004, reggaeton became popular in the United States and Europe, becoming a genre with great presence in Billboard Top Latin Albums charts in 2006 [56].

2.5.7 The 2010s

The 2010s continued with the popularity of pop music and reggaeton, but this decade has been also characterized by the emergence of a new hip hop subgenre: trap. During the early-to-mid 2000s, trap music began to emerge as a recognized genre after the mainstream success of a number of albums and singles with lyrics that covered topics about life in *the trap*, drug dealing and the struggle for success. Trap music genre (term originated in Atlanta) is a hip hop subgenre originated during the 1990s from Southern hip hop in the Southern United States. It is typified by its ominous lyrics and sound that incorporate a heavy use of multi-layered hard-lined and melodic synthesizers; crisp, grimy, and rhythmic snares; deep 808 kick drums; double-time, triple-time and similarly divided hi-hats; and a cinematic and symphonic utilization of string, brass, woodwind, and keyboard instruments creating an overall dark, harsh, grim, and bleak atmosphere for the listener [57, 44].

First it could be noticed the trap influence in mainstream songs around 2006. Secondly, by the end of the 2000s decade, a second wave of trap artists continued to gain momentum and frequently top the Billboard hip hop charts. Since trap producer Lex Luger's rise -he went on to produce more than 200 songs between 2010 and 2011-, many of his sounds have since been adopted and incorporated by other hip hop producers, trying to replicate his success, as Luger is often credited with popularizing the modern trap sound [57]. Since maintaining a strong presence on the mainstream music charts, trap music has been utilized by non-hip hop artists. In 2016 and 2017 trap-influenced songs and trap songs began their biggest commercial success appearing frequently in first position of Billboard Hot 100 chart.

2.5.8 Main revolutions of Western popular music since 1950s

After this historical review of Western popular music from 1950s to actuality, we can suppose the apparent influence to mainstream music of the following popular music revolutions:

- 1951: rock and roll emergence
- 1964: British pop emergence
- 1977: disco popularity spreadness across USA
- 1977: punk emergence (especially in UK)
- 1991-1992: hip hop's mainstream emergence
- 1991: grunge success
- 1994: punk rock resurgence in USA
- 1997-2000: boy bands, girl bands and pop starlets emergence
- 2006: reggaeton mainstream emergence
- 2006: first trap mainstream influence
- 2016-2017: second wave of trap

In next chapter the selection of some of these revolutions will be discussed in order to define the analysed years of the study.

Chapter 3

Methodology

This chapter exposes the procedures followed to design and implement the data-set and the system able to describe music timbre of Billboard Hot 100 along last decades. State of the art about history of Western popular music and previous timbre studies, characterizations and classification models will guide the orientation and the sense of developing data-set and timbre characterization system. Figure 3.1 shows the steps followed in the methodology.

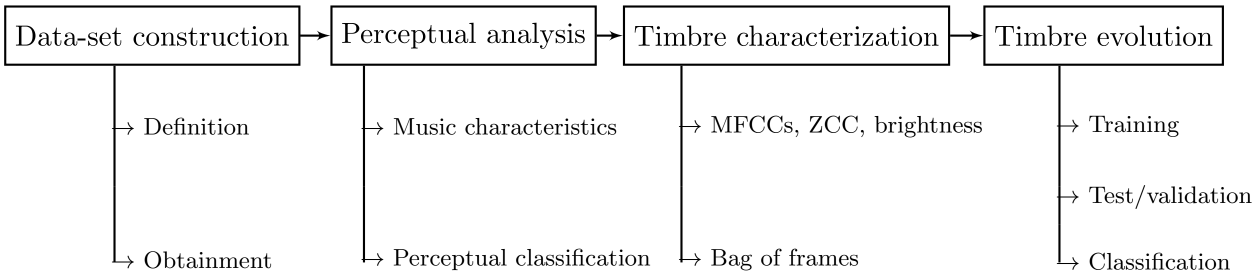


Figure 3.1: Scheme of the developed methodology

3.1 Data-set

The goal designing the data-set of the project is to use a database capable to represent in mainstream music several Western popular music revolutions -and a sample of mainstream music before and after these moments- in order to achieve the targets of the project.

For this purpose all songs of each studied year which reached the Billboard Hot 100 first position every week [2] (often the same song maintains the first position several weeks) have been selected. So, despite of consideration Billboard Hot 100 music chart as representative of mainstream music in worldwide terms, it is important to take into account that study is focused in US mainstream music.

Due to the limitation of Billboard Hot 100 music chart existence (since 1958) and the US and mainstream audience characterization of this chart, there were selected six of the music revolutions presented in section 2.5.8 in order to build the data-set. Every selected music revolution defines a group of 5 consecutive years (around the concrete year selected before) which includes all songs which reached the first position of Billboard Hot 100 some week in those years. These are the five groups of the data-set:

- 1962-1966 - British pop emergence (119 songs)
- 1975-1979 - Disco and punk popularity (134 songs)
- 1989-1993 - Hip hop's mainstream emergence (110 songs)
- 1997-2001 - Boy bands, girl bands and pop starlets emergence (74 songs)

- 2012-2016 - Second wave of trap (54 songs)

The data-set is made up of 492 songs in Waveform Audio File (WAV) format. All files are audio lossless compressed and sampled at 44100 Hz with 16 bits. Appendix C shows the tables with all constructed data-set and its perceptual classification.

3.2 Perceptual analysis

Once obtained all songs of data-set the next step consists in analyse them perceptually. In order to detect and identify some timbral changes that can be related to music revolutions defined, the author of this study defined and tried to listen -or not to- several music characteristics in every 5-years-group of songs (see appendix C). So perceptual analysis consists in a music characteristics definition and the perceptual classification (yes/no) of data-set for each of them. These music characteristics are described as follows.

3.2.1 1962-1966 - British pop emergence

- **Vocal accompaniment (choruses) in close harmony**: it refers to those singer voices which are not a lyric contribution to the song, are backing vocals, its semantic content is different from principal voice -resembling doo-wop, usually with nonsense syllables- and are in close harmony to the song [39]. This is a very characteristic distinction of Merseybeat groups (see section 2.5.2), including The Beatles, so could characterize the British Invasion.

Table C.1 in appendix C shows this perceptual classification for all 1962-1966 songs data-set. For example, The Beatles *A Hard Day's Night* and *I Feel Fine* songs have got the presence of vocal accompaniments in close harmony according to the perceptual listening. Instead, The Crystals *He's A Rebel* and The Tornados *Telstar* songs have not got this characteristic.

3.2.2 1975-1979 - Disco and punk popularity

- **Backing keyboard instruments (piano, electric organ, string synth and electromechanical keyboards)**: it is a very characteristic component of disco music [53], present in both choruses and verses (see section 2.5.3). Its identification corresponds to instrument recognition, a task only and clearly related to timbre.
- **Orchestral instruments (bowed string or brass sections)**: this characteristic refers to both bowed string and brass instruments in a non-exclusive way. It is also a component of disco music (see section 2.5.3) and an instrument recognition task.
- **Distorted power chords**: this sonority majority produced by guitar distortion was introduced for the first time by punk genre and it is its distinctive feature (see section 2.5.3).

3.2.3 1989-1993 - Hip hop's mainstream emergence

- **Drum machine**: it is related to hip hop's evolution occurred around the time that sampling technology and drum-machines became widely available to the general public at a cost that was affordable to the average consumer (see section 2.5.4). So it tries to distinguish between those backing instrumentals produced by drum machine -electronic sound- from those interpreted by musicians with organic instruments.
- **High load of low frequencies**: a subjective characteristic interpreted by the author according to hip hop songs features. It tries to discriminate those songs with constant load of low sounds produced by low voices or kick drums.

3.2.4 1997-2001 - Boy bands, girl bands and pop starlets emergence

- **Electronic backing**: it corresponds to backing instrumentals produced with electronic sounds, due to the incorporation of new synthesizers in music production.
- **Vocal accompaniment (choruses) in close harmony**: same as section 3.2.1.

3.2.5 2012-2016 - Second wave of trap

- **Sub-bass double/triple kick drums**: this is a very characteristic sonority of trap, loading low frequencies, produced commonly by 808 drum machine (see section 2.5.7).
- **Double-time, triple-time, twitchy hi-hats**: it is also a distinctive feature of trap songs, referring to high frequencies in this case (see section 2.5.7).

Table C.5 shows this perceptual classification for all 2012-2016 songs data-set. For example, Desiigner *Panda* song has got the presence of sub-bass double/triple kick drums and double-time, triple-time, twitchy hi-hats, while *Hello* by Adele has not got these characteristics. And *What Do You Mean* by Justin Bieber has got the presence of sub-bass double/triple kick drums but it has not got the presence of double-time, triple-time, twitchy hi-hats.

3.3 Timbre characterization

Once data-set was constructed and classified perceptually, next step was to implement a system able to characterize timbre of all these songs, specially by the aspects defined in section 3.2. In order to achieve this, MFCCs (section 2.3.1) were computed using a function of MIRtoolbox (1.6.1 version), a Matlab toolbox [29]. Using the same toolbox also ZCC and brightness (amount of spectral energy corresponding to frequencies higher than spectral centroid frequency approximation) were computed.

Firstly, the system reads all song files through *csv* tables containing all information (file name, weeks on Hot 100's first position and perceptual classification). Audio files are opened with sampling rate of 44100 Hz and 16 bits, dividing them into frames of 50ms and half

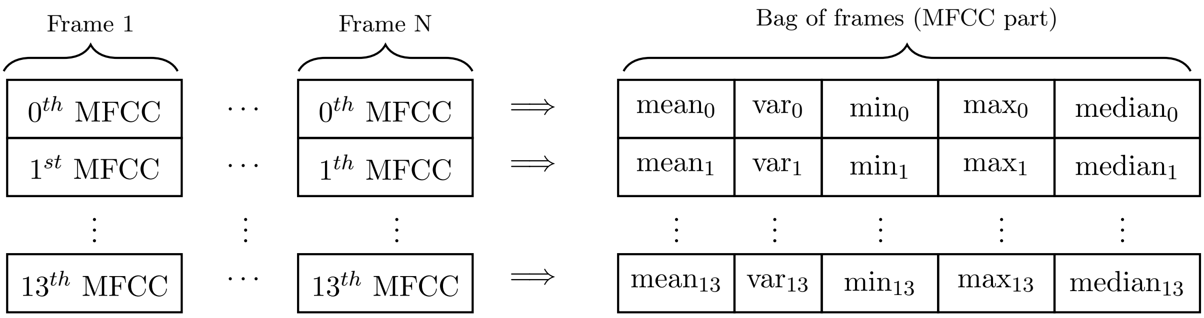


Figure 3.2: BOF construction from computed MFCCs. Complete BOF also contains the features obtained from deltas, ZCC and brightness. The figure shows the procedure for one song file.

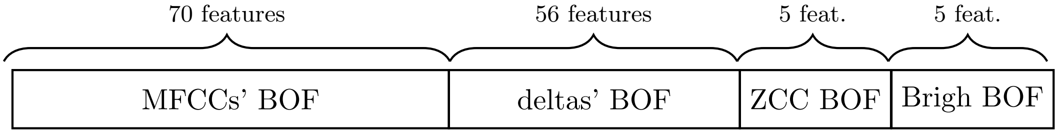


Figure 3.3: Features vector of each song

overlapping (each frame begins at the middle of the previous frame). Then, with Mel-band spectrum decomposed in 40 bands, the first 14 MFCCs are computed (including the 0^{th} component) and additionally one delta-MFCC and one delta-delta-MFCC, calculated as the difference between any two consecutive values of 14 MFCC components and between any two consecutive values of delta-MFCC components, respectively (deltas function by Dan Ellis was used for this purpose [11]).

Finally, in order to obtain the bag-of-frames (BOF) -a group of measures able to represent each coefficient of all song frames independently from time-, several statistics were computed for each MFCC component and for two delta components. Mean, variance, minimum, maximum and median functions were calculated across all song frames for each of the 14 MFCC components, obtaining 14x5 values for each song (see figure 3.2). Mean and variance functions were computed across all song frames for delta-MFCC and delta-delta-MFCC, obtaining 14x4 values for each song. Totally, each song file is described by 9 statistics across 14 components, together with 10 more components of ZCC and brightness, obtaining vectors of 136 features (see figure 3.3).

ZCC is computed using the same method to open audio files and divide them into frames of 50ms and half overlapping. In this case the BOF is composed by mean, variance, minimum, maximum and median functions calculated for ZCC values of all frames, obtaining 5 more features for each song.

Brightness is calculated as the amount of spectral energy corresponding to frequencies higher than a given cut-off threshold, defined at 1500Hz [26]. Frames are obtained with the same window length and overlapping. BOF is composed by the same functions as ZCC BOF, obtaining 5 more features for each song.

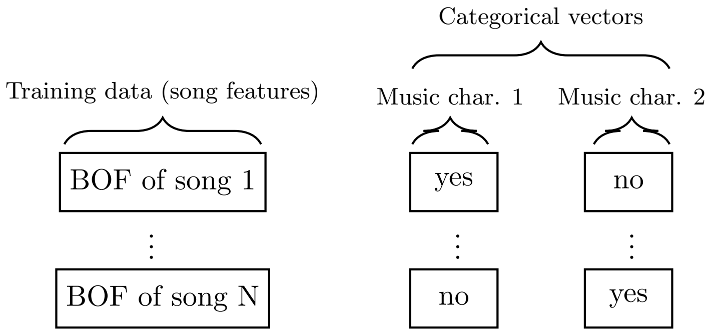


Figure 3.4: Scheme of training data

3.4 Timbre evolution

With BOF of all songs data-set as features, the system uses SVMs in order to obtain models able to classify timbre blend into binary classes defined at section 3.2 (music characteristics). First of all, the system has to be trained with labeled data (perceptual classification) and then the obtained SVM models have to be tested and validated with evaluation measures. Finally, these models will classify data-sets into defined classes, tracing *a timbre evolution*.

3.4.1 Training

The procedure followed to train the SVM classifier consists in implement a SVM MATLAB function with ground truth perceptually developed: a matrix of training data and a categorical vector with each row representing a class label. Matrix of training data is composed by rows which correspond to observations or replicates (songs) and columns which corresponds to features (136 at most). The categorical vector corresponds to the perceptual classification (yes/no) of each song for one of the music characteristics (see section 3.2). In this way, if SVM classifier is trained for every music characteristic, it can be obtained one SVM model for every of them. Figure 3.4 shows this explanation.

It is important to remember that data-set is divided into five 5-years-group (see section 3.1), with different music characteristics perceptually labeled in them (according to the music historical moment). So every SVM classifier is trained with categorical vector of one music characteristic and its associated training data of the 5-years-group which is classified in this characteristic.

Another important fact to take into account is the training data's number of features and different SVM parameters able to be configured. A large number of features could overfitting the system while using only a few features could produce underfitting (see section 2.4.2). Different kernel functions also can be applied to SVM training system in order to automatically realize a non-linear mapping to a feature space if data is not linearly separable (see section 2.4.1). All configurations used for each music characteristic training are exposed with their results in chapter 4.

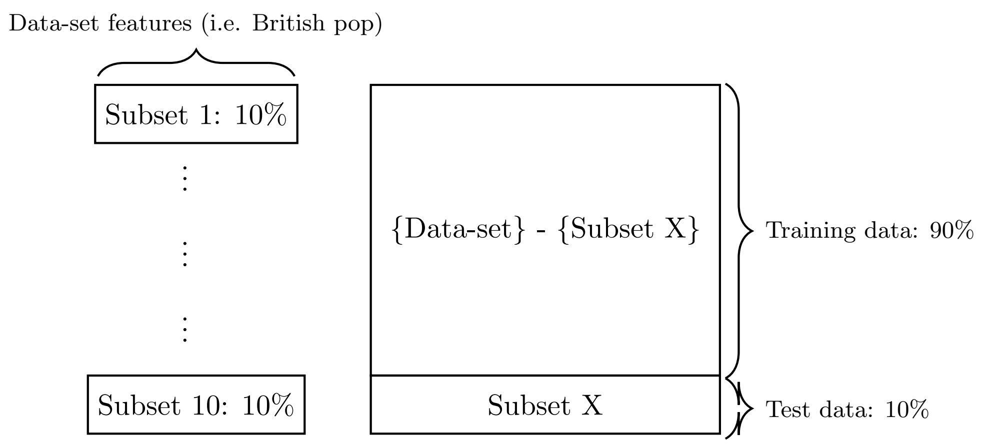


Figure 3.5: Scheme of SVMs 10-fold cross-validation process. It has to be iterated 10 times for X=1:10

3.4.2 Test/validation

Two methods were used to divide the ground truth data into training and test or validation data. The first one uses, for every music characteristic, 80% of corresponding data-set to train the SVM system and the remaining 20% to test it. Train and test accuracies are computed. And the second one is a 10-fold cross-validation. It uses 10 times, with different subset each iteration, 90% of data-set for training and 10% for testing (see figure 3.5). Then, the evaluation measure are obtained as mean of measures of every iteration.

The reason to use the second method is the lack of data (the smallest data-set has only 54 songs). In this way it is possible to obtain more reliable validations of the SVMs.

3.4.3 Timbre evolution prediction

The final process consists in measuring the evolution of timbre in Billboard Hot 100. For this purpose, with the selected timbre features (MFCCs, ZCC and brightness), it is possible to classify all data-sets (five 5-years-groups) into binary classes (yes/no) for all music characteristics (there are 10). Using the trained and tested SVMs for each of these music topics with all unlabeled data (all data-set songs) will show the evolution of these 10 music characteristics along time.

Chapter 4

Results

This chapter exposes the results obtained after following the methodology presented in chapter 3.

4.1 Timbre characterization

With MFCCs, ZCC and brightness features (see section 2.3) and configurations defined at section 3.3, the computations for all data-set have been obtained. They are illustrated with two different examples at appendix D. First one is The 4 Seasons' *Sherry*, one of the first position hit of Billboard Hot 100 in 1962, during the period of 1962-1966 data-set group. And the second one is Snow's *Informer*, one of the first position hit of Billboard Hot 100 in 1993, during the period of 1989-1993 data-set group.

Notice that higher values for ZCC and brightness in the second song indicate more presence of percussive sounds and more amount of energy in high frequencies, respectively. It seems logic because the second song has an aggressive percussion instrumental with high-pitched strident electronic sounds.

4.2 Evaluation of the timbre characteristics classifiers

This section exposes the used configurations of the SVM training system, and its achieved results, in order to obtain the best measures. Accuracy, precision, recall and F-measure were selected as these measures because of their good performance for statistical analysis and evaluation of binary classification (see appendix E). The values of these measures are between 0 and 1: while values close to 0 are poor classifications, values close to 1 represents a good classification.

Training and validation results exposed here correspond only to the 10-fold cross-validation test method, explained at section 3.4.2. So the measures took into account are means of all iterations (including F-measure), and sometimes they can not be computed (represented with '-' in tables) because some iteration has classified all data (a few songs) at the same class with errors (and in this way it is not possible to compute precision and/or recall in those iterations, so the mean gives can not be computed). Also it is possible that in some iteration precision and recall are 0 so F-measure can not be computed.

All training SVM systems were configured firstly with the same configuration: configuration 1. It corresponds to the use of 126 features (MFCCs and deltas) with a linear kernel. Then, a second configuration, implemented according to music characteristic and its features, is exposed in order to show improvements or impairments of the system. Finally, after several configurations not exposed, the selected SVM system configuration for each of the 10 music characteristics is the last one of each of the following tables: configuration 3. These configurations have obtained the best measures and have been used to classify all data-set in section 4.3.

4.2.1 Vocal accompaniment (choruses) characteristic

Configuration 2 corresponds to the use of 126 MFCCs features plus 10 features corresponding to brightness and ZCC BOF. Kernel used is Gaussian Radial Basis Function (RBF) to map the training data into kernel space.

Configuration 3 corresponds to the use of variances of delta-MFCCs and delta-delta-MFCCs, variance, minimum and median of MFCC coefficients and mean, variance, minimum and maximum statistics of ZCC and brightness values. In totally 78 features are used. Kernel function used is the polynomial.

	Training				**Test**				Feat.	Ker.
	Ac.	Pre.	Rec.	F.	Ac.	Pre.	Rec.	F.		
Config. 1	1	1	1	1	0.536	0.544	0.633	0.573	126	linear
Config. 2	1	1	1	1	0.420	-	0.414	-	136	rbf
Config. 3	1	1	1	1	0.664	0.706	0.674	0.658	78	poly.

Table 4.1: Training and test measures of vocal accompaniment characteristic for different configurations of 1962-1966 data-set

4.2.2 Backing keyboard instruments characteristic

Configuration 2 corresponds to the use of 126 MFCCs features plus 5 features corresponding to brightness BOF. Kernel used is the polynomial.

Configuration 3 corresponds to the use of variances of delta-MFCCs and delta-delta-MFCCs, mean, variance and median of MFCC coefficients and mean, variance, minimum and maximum statistics of ZCC values. In totally 74 features are used. Kernel function used is the quadratic.

	Training				**Test**				Feat.	Ker.
	Ac.	Pre.	Rec.	F.	Ac.	Pre.	Rec.	F.		
Config. 1	0.996	0.998	0.997	0.997	0.631	0.735	0.704	0.706	126	linear
Config. 2	0.985	0.988	0.990	0.989	0.633	0.704	0.842	0.751	131	poly.
Config. 3	1	1	1	1	0.704	0.739	0.876	0.790	74	quadr.

Table 4.2: Training and test measures of backing keyboard characteristic for different configurations of 1975-1979 data-set

4.2.3 Orchestral instruments characteristic

Configuration 2 corresponds to the use of means of delta-MFCCs and delta-delta-MFCCs, mean, variance, minimum and median of MFCC coefficients and 5 features corresponding to brightness BOF. Kernel used is the quadratic.

Configuration 3 corresponds to the use of variances of delta-MFCCs and delta-delta-MFCCs, and mean, variance and median of MFCC coefficients. In totally 70 features are used. Kernel function used is the quadratic.

	Training				**Test**				Feat.	Ker.
	Ac.	Pre.	Rec.	F.	Ac.	Pre.	Rec.	F.		
Config. 1	0.999	1	0.999	0.999	0.584	0.734	0.729	0.688	126	linear
Config. 2	1	1	1	1	0.640	0.749	0.816	0.741	89	quadr.
Config. 3	1	1	1	1	0.689	0.755	0.862	0.776	70	quadr.

Table 4.3: Training and test measures of orchestral instruments characteristic for different configurations of 1975-1979 data-set

4.2.4 Distorted power chords characteristic

Configuration 2 corresponds to the use of means and variances of delta-MFCCs and delta-delta-MFCCs, and mean and variance of MFCC coefficients. Kernel used is the linear.

Configuration 3 corresponds to the use of means of delta-MFCCs and delta-delta-MFCCs, median of MFCC coefficients and 5 features corresponding to brightness BOF. In totally 47 features are used. Kernel function used is the polynomial.

	Training				**Test**				Feat.	Ker.
	Ac.	Pre.	Rec.	F.	Ac.	Pre.	Rec.	F.		
Config. 1	1	1	1	1	0.914	-	-	-	126	linear
Config. 2	1	1	1	1	0.921	-	-	-	84	linear
Config. 3	1	1	1	1	0.928	-	-	-	47	poly.

Table 4.4: Training and test measures of distorted characteristic for different configurations of 1975-1979 data-set

4.2.5 Drum machine characteristic

Configuration 2 corresponds to the use of means and variances of delta-MFCCs and delta-delta-MFCCs, mean, variance, minimum, maximum and median of MFCC coefficients, and 10 features corresponding to ZCC and brightness BOF. Kernel used is the linear one.

Configuration 3 corresponds to the use of means of delta-MFCCs and delta-delta-MFCCs, median of MFCC coefficients and 10 features corresponding to ZCC and brightness BOF. In totally 52 features are used. Kernel function used is the polynomial.

	Training				Test				Feat.	Ker.
	Ac.	Pre.	Rec.	F.	Ac.	Pre.	Rec.	F.		
Config. 1	1	1	1	1	0.700	0.607	0.677	0.602	126	linear
Config. 2	1	1	1	1	0.700	0.607	0.677	0.602	136	linear
Config. 3	1	1	1	1	0.773	0.860	0.577	0.624	52	poly.

Table 4.5: Training and test measures of drum machine characteristic for different configurations of 1989-1993 data-set

4.2.6 High load of low frequencies characteristic

Configuration 2 corresponds to the use of variances of delta-MFCCs and delta-delta-MFCCs, and variance, minimum, maximum and median of MFCC coefficients. Kernel used is the polynomial.

Configuration 3 corresponds to the use of means of delta-MFCCs and delta-delta-MFCCs, and mean, variance and median of MFCC coefficients. In totally 70 features are used. Kernel function used is the polynomial.

	Training				Test				Feat.	Ker.
	Ac.	Pre.	Rec.	F.	Ac.	Pre.	Rec.	F.		
Config. 1	1	1	1	1	0.764	0.577	-	-	126	linear
Config. 2	1	1	1	1	0.764	0.577	-	-	84	poly.
Config. 3	1	1	1	1	0.791	0.750	-	-	70	poly.

Table 4.6: Training and test measures of high load of low frequencies characteristic for different configurations of 1989-1993 data-set

4.2.7 Electronic backing characteristic

Configuration 2 corresponds to the use of means of delta-MFCCs, mean, variance and median of MFCC coefficients and 5 features corresponding to the ZCC BOF. Kernel used is the quadratic.

Configuration 3 corresponds to the use of 5 features of ZCC BOF. Kernel function used is the quadratic.

	Training				Test				Feat.	Ker.
	Ac.	Pre.	Rec.	F.	Ac.	Pre.	Rec.	F.		
Config. 1	1	1	1	1	0.714	0.785	0.821	0.786	126	linear
Config. 2	1	1	1	1	0.623	0.691	0.704	-	61	quadr.
Config. 3	0.903	0.926	0.933	0.930	0.831	0.890	0.815	0.839	5	quadr.

Table 4.7: Training and test measures of electronic backing characteristic for different configurations of 1997-2001 data-set

4.2.8 Late 1990s vocal accompaniment (choruses) characteristic

Configuration 2 corresponds to the use of means of delta-MFCCs and delta-delta-MFCCs, mean, variance and median of MFCC coefficients, and 5 features corresponding to ZCC BOF. Kernel used is the linear.

Configuration 3 corresponds to the use of means of delta-MFCCs and delta-delta-MFCCs, mean of MFCCs coefficients and 10 features corresponding to ZCC and brightness BOF. In totally 52 features are used. Kernel function used is the linear.

	Training				**Test**				Feat.	Ker.
	Ac.	Pre.	Rec.	F.	Ac.	Pre.	Rec.	F.		
Config. 1	1	1	1	1	0.575	-	0.553	-	126	linear
Config. 2	0.906	0.896	0.995	0.932	0.701	0.655	0.779	-	75	linear
Config. 3	0.967	0.965	0.990	0.974	0.740	0.700	0.806	-	52	linear

Table 4.8: Training and test measures of late 1990s vocal accompaniment characteristic for different configurations of 1997-2001 data-set

4.2.9 Sub-bass kick drums characteristic

Configuration 2 corresponds to the use of variances of delta-MFCCs and delta-delta-MFCCs, and variance of MFCC coefficients. Kernel used is the polynomial.

Configuration 3 corresponds to the use of mean and median of MFCCs coefficients and 5 features corresponding to ZCC BOF. In totally 33 features are used. Kernel function used is the polynomial.

	Training				**Test**				Feat.	Ker.
	Ac.	Pre.	Rec.	F.	Ac.	Pre.	Rec.	F.		
Config. 1	1	1	1	1	0.650	-	-	-	126	linear
Config. 2	0.998	1	0.994	0.997	0.691	-	-	-	42	poly.
Config. 3	1	1	1	1	0.759	-	-	-	33	poly.

Table 4.9: Training and test measures of sub-bass kick drums characteristic for different configurations of 2012-2016 data-set

4.2.10 Twitchy hi-hats characteristic

Configuration 2 corresponds to the use of means of delta-MFCCs and delta-delta-MFCCs, and variance, minimum, maximum and median of MFCC coefficients. Kernel used is the polynomial.

Configuration 3 corresponds to the use of mean of delta-delta-MFCCs, median of MFCCs coefficients and 5 features corresponding to ZCC BOF. In totally 33 features are used. Kernel function used is the polynomial.

	Training				Test				Feat.	Ker.
	Ac.	Pre.	Rec.	F.	Ac.	Pre.	Rec.	F.		
Config. 1	1	1	1	1	0.777	-	-	-	126	linear
Config. 2	1	1	1	1	0.714	-	-	-	84	poly.
Config. 3	1	1	1	1	0.786	-	-	-	33	poly.

Table 4.10: Training and test measures of twitchy hi-hats characteristic for different configurations of 2012-2016 data-set

4.2.11 Best configurations for the music characteristics

Configurations with the best test measures (best F-measure if there is or best accuracy otherwise) for all music characteristics presented before are resumed in next table:

Music char.	Config.	Ac.	Pre.	Rec.	F.	Feat.	Ker.
Vocal accompaniment	3	0.664	0.706	0.674	0.658	78	poly.
Backing keyboard	3	0.704	0.739	0.876	0.790	74	quadr.
Orchestral	3	0.689	0.755	0.862	0.776	70	quadr.
Distorted chords	3	0.928	-	-	-	47	poly.
Drum machine	3	0.773	0.860	0.577	0.624	52	poly.
Low frequencies	3	0.791	0.750	-	-	70	poly.
Electronic backing	3	0.831	0.890	0.815	0.839	5	quadr.
90s vocal accompaniment	3	0.740	0.700	0.806	-	52	linear
Sub-bass kick drums	3	0.759	-	-	-	33	poly.
Twitchy hi-hats	3	0.786	-	-	-	33	poly.

Table 4.11: Best configurations measures for the music characteristics

4.3 Classification and timbre evolution

Next five figures show the evolution in Billboard Hot 100 of the 10 music characteristics evaluated in this project. SVM models obtained before, which train and test processes have been presented in section 4.2, have classified all music data-sets (five groups of 5 years) into these classes (if songs contain or not the music characteristics). The evolution of the characteristics is measured according to a percentage of presence in Billboard Hot 100's first position every analysed year. This year percentage is computed by weighting each song which has been classified positively with the number of weeks that had been on Hot 100's first position, and dividing the sum of weights by the number of songs of that year.

Discussion and conclusion of these results are in section 4.4.

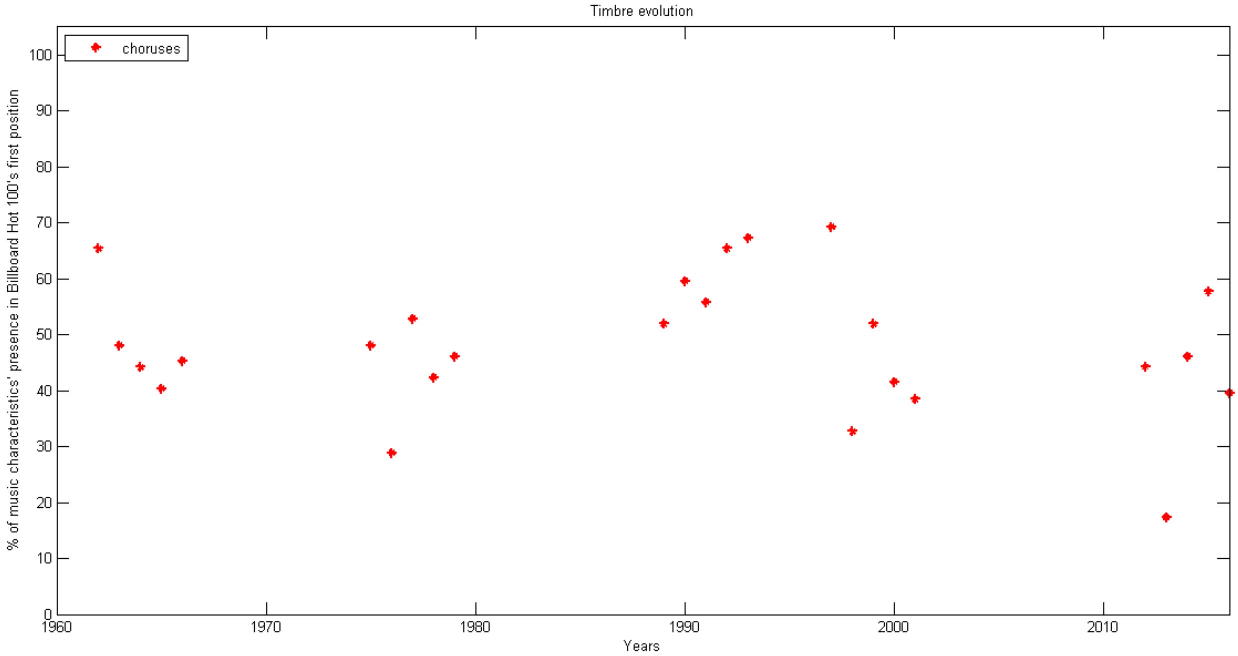


Figure 4.1: Vocal accompaniment (choruses) evolution in Billboard Hot 100

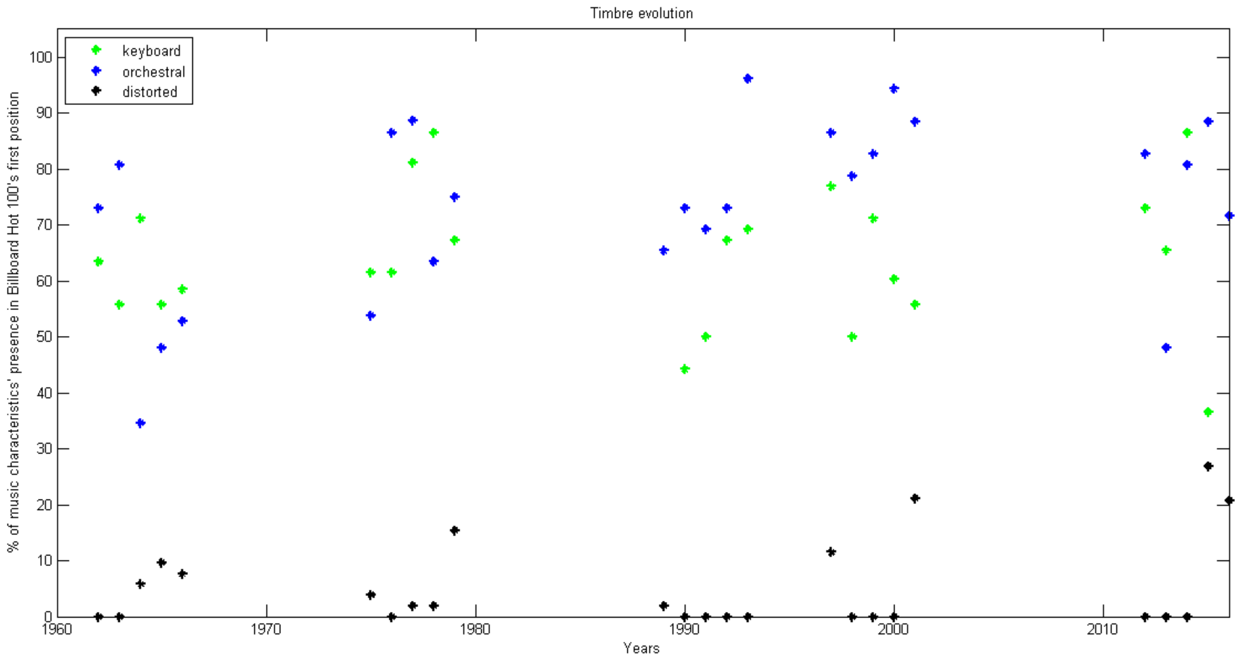


Figure 4.2: Backing keyboard instruments, orchestral instruments and distorted power chords evolution in Billboard Hot 100

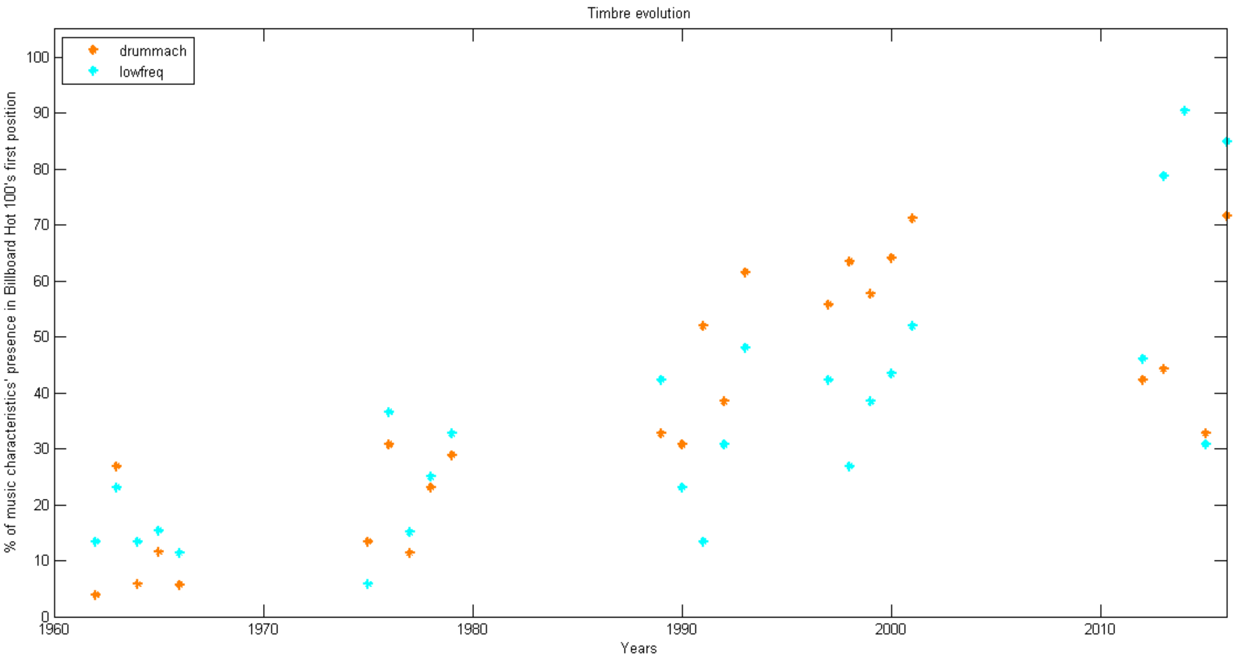


Figure 4.3: Drum machine and high load of low frequencies evolution in Billboard Hot 100

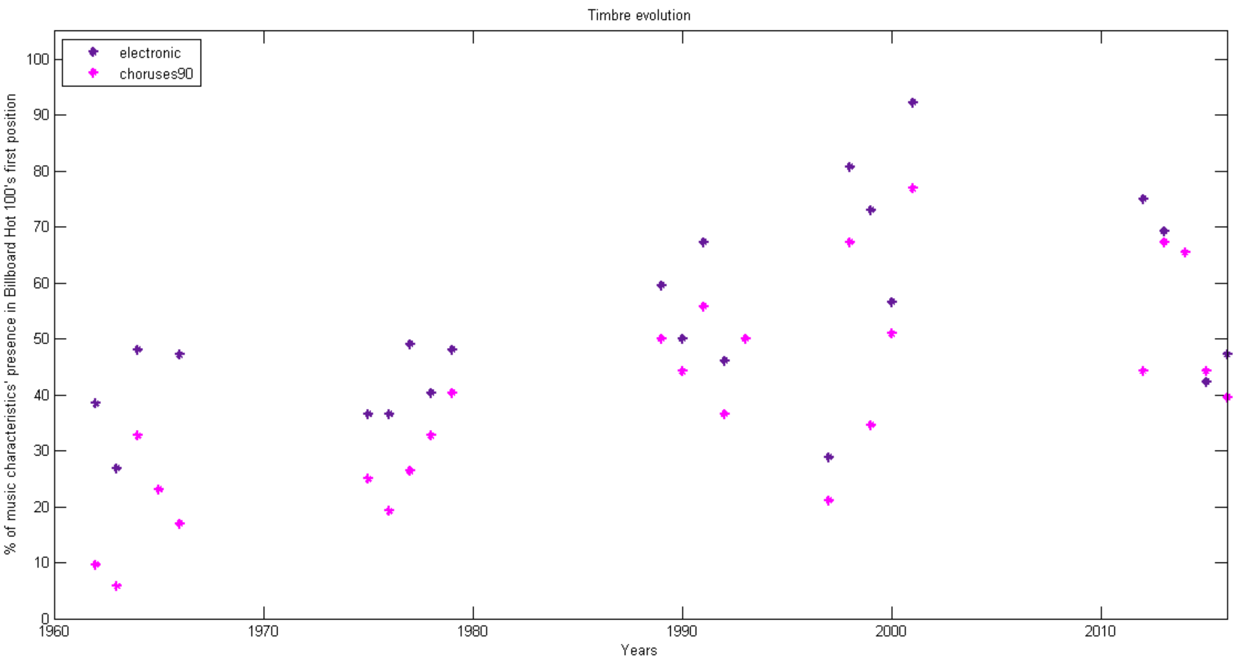


Figure 4.4: Electronic backing and late 1990s vocal accompainment (choruses) evolution in Billboard Hot 100

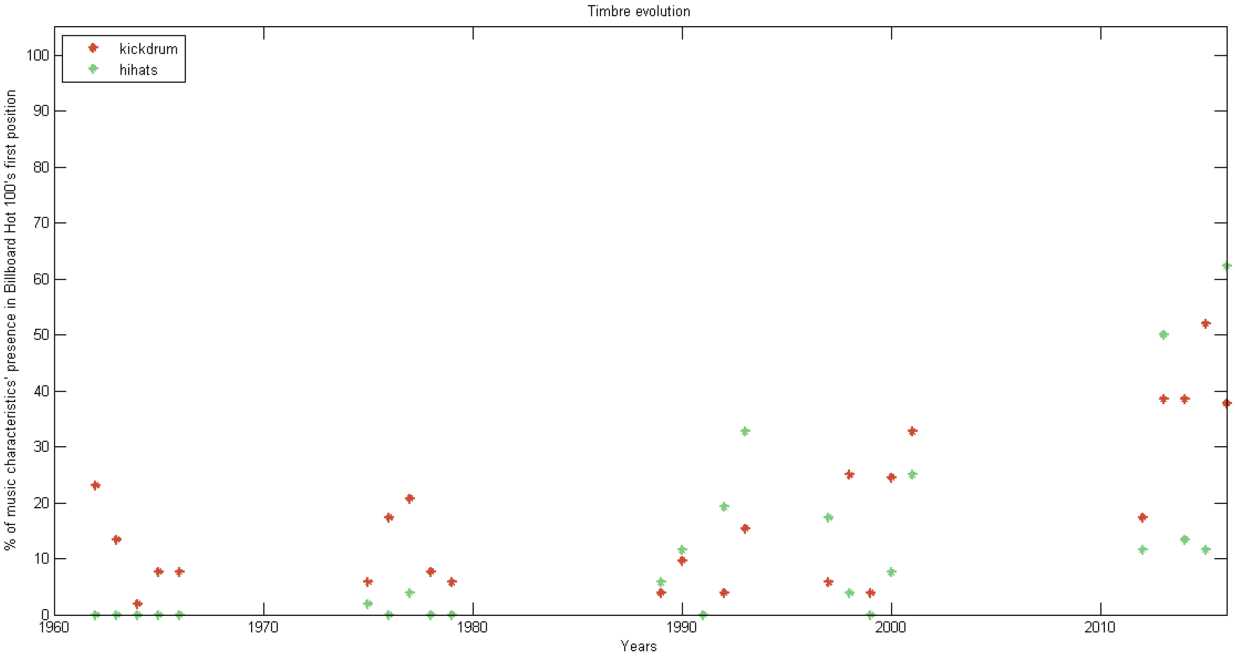


Figure 4.5: Sub-bass kick drums and twitchy hi-hats evolution in Billboard Hot 100

4.4 Discussion of the results

Vocal accompaniment evolution (figure 4.1) not seems to show any recognizable pattern but it can be for many reasons. This characteristic has a mean presence around 50% in all analysed groups of years except 1990s group, with a little bit more presence. This fact may be due to two possibilities: or system has not performed well and the characterized timbre feature is not correct, or it is an indicator that this music characteristic is not representative of British pop invasion in 1960s. Low test measures (the lowest) exposed at section 4.2.1 strengthen the first possibility, while the presence of this characteristic in 1990s group can reinforce the second one. In any case it is a difficult challenge to define and characterize a feature which presence is punctual and has a lot of timbral content (vocal choruses).

Disco and punk characteristics -backing keyboard and orchestral instruments, and distorted power chords- show a few more things (figure 4.2). Keyboards and orchestral instruments have a big presence in all decades (>50%), and it seems to be logic according to the extended use of these instruments in all genres and epochs. But these characteristics had supposed to define disco genre, and although the biggest presence of backing keyboards is in 1970s group, orchestral instruments have more presence in late 1900s and 2000s. The reason that explains this could be outside the field of timbre study because disco keyboards have certain amount of characteristics such as the moment when are played, the rhythm, the loudness, the interpreted melodies or harmonies, etc. That is to say, one instrument can be played in a specific way and determine a genre characteristic but it also can be played in a different way (obviously keeping the same timbre). Referring to distorted power chords the results show a very little presence in all years (several years is 0%) and it has sense. Although it is a punk defining feature, very few songs were perceptually classified in it in

19070s group, so it can be affirmed that punk has not influenced Billboard Hot 100.

Drum machine and high load of low frequencies evolution (figure 4.3) shows clearly an increase of their presence in Billboard Hot 100. It is very logic that drum machine use has increased together with electronic advances, digitalization and its associated improvement (notice it in 1990s, 2000s and 2010s). But ideally, it should not has presence in 1960s and 1970s (when drum machines were not used), so its low percentages in those years can be attributed to classification errors. In the same sense, it seems easily recognizable that high load of low frequencies has more presence in electronic and modern genres such hip hop or trap (surely sub-bass kick drum characteristic has contributed). It can be said that mainstream music is getting more and more high load of low frequencies. Also it seems to be a correlation between these two characteristics and it may be because the utilization of drum machines usually go related to its monotonous and continuous use (high load of low frequencies).

Electronic backing and vocal accompaniment of late 1990s evolution (figure 4.4) shows principally the difference of electronic backing before (<50%) and after late 1980s (>50%) with the hip hop genre rise. The reasons may be the same that explained before for drum machine case. So another conclusion is that Billboard Hot 100 is evolving to more electronic sounds. Increase of vocal accompaniment presence not seems to have any explanation beyond a bad classification or a bad timbre characterization.

Finally, trap characteristics -sub-bass kick drums and twitchy hi-hats- present a clearly evolution along time (4.5). As they are very exclusive characteristics of trap, they obtain the biggest presence in 2010s, but it can be observed a notably presence of twitchy hi-hats also in late 1980s and 1990s. This is because rap songs (hip hop songs) also usually has drum machines or sounds with big presence of hi-hats (the difference is in the rhythm). So with regard to musical timbre, hip hop and trap have influenced Billboard Hot 100, or they have become mainstream music in some cases at least. Sub-bass kick drum characteristic not seems to be well characterized because this kind of kick (its low frequencies) has not been listened in perceptual analysis before 2010s. But it can be interpreted as the recognition of normal kick drums so the evolution would show a bigger utilization of kick drums since late 1990s.

Chapter 5

Budget

This project has been developed using free resources available on Internet (all bibliography) and MATLAB software, free for UPC students [7]. Also MIRTOOLBOX was used, a free-available MATLAB toolbox [29].

Accordingly, the estimated budget of the study consists basically in costs associated to the salary of the researches, the time spent in it and the hardware used.

The compute of the salaries has been done considering the dedication in the project of one junior engineer (the author of the work) and two senior engineers (supervisors of the work).

The duration of the project was 22 weeks, as considered in Gantt diagram (see section 1.4.2).

	Amount	**Wage/hour**	**Dedication**	**Total**
Junior engineer	1	15.00 €/h	40 h/week	13,200 €
Senior engineer	1	25.00 €/h	1 h/week	550 €
Senior engineer	1	35.00 €/h	0.5 h/week	385 €
			Total	14,135 €

Table 5.1: Estimated budget of the project

Chapter 6

Conclusions

This study has focused on musical timbre complexities and it has tried to characterize and classify the timbre blend of some Billboard Hot 100's first position songs in last decades. The objective was to trace the timbre evolution of mainstream music, and to see its relation to Western popular music.

For this purpose, first of all, timbre study bibliography has been analysed focusing in two points of view depending on the discipline which study it: MIR and music psychology. Then, several music descriptors have been studied and defined in order to characterize timbre blend: MFCCs, ZCC and brightness. Also, a song data-set with its partial perceptual classification (or labeling) was designed and obtained with lossless quality, as mainstream music representation. To achieves this, five periods of five years and ten music characteristics associated were defined, dividing the data-set into five 5-years-groups. After that, one SVM model for each music characteristic has been implemented and configured with accuracy results over 65% for all cases, and F-measures over 62% (except some models which F-measure could not been computed). Finally, these models have classified all songs of the data-set, showing an evolution of the timbre characteristics in Billboard Hot 100 along last decades. Despite of the low veracity of the results, they have allowed to discuss them according to Western popular music evolution and to see the shortcomings of the classifiers. The project also has showed the difficulties of characterizing timbre and the possible potential of this audio feature to describe music and its evolution.

The intention of this project was to describe mainstream music evolution in timbre blend terms, and its relation to Western popular music evolution. This means that ideal results would show how the perceptual sonority, the timbre blend, evolves and changes along time, in determinate years, influenced by new sounds, technologies and genres that become fashionable to mass audiences.

Maybe the structure of the project, its methodology, the objectives and goals, and the knowledge that is pretended to reach are more robust than the obtained results or their validity. But this is an assumed risk when we try to characterize complex things such as the musical timbre or cultural expressions and their evolution. And the robustness of the bases of the project is the warranty that it can be improved or redefined in the future in regard to results (timbre classification), timbre descriptors, classification model and data-set used. The most important is not to fall into false attributions, so discussion and conclusions has to be wise with generalizations or causality relations. For example, the rise of electronic backing characteristic usage can not be assumed immediately as the rise of hip hop genre, because there are more genres which have this feature. But the rise together of electronic backing, drum machine usage and high load of low frequencies could be a much better indicator of the hip hop emergence. And even so, it is risky and could be wrong to make these restricted attributions from a musicological point of view. In this sense, timbre evolution results presented at section 4.3 can be interpreted in many ways, and their discussion (see section 4.4) has been careful with these aspects.

The principal handicap of this study has been the lack of data. It is very difficult to develop and implement good classification systems with data-sets of 50 or 60 observations (labeled data) and 491 expectations (unlabeled data). And it has to be added the impossibility to make reliable or valid conclusions of timbre evolution from a musicological point of view with a low percentage of samples (the data-set used could represent only a very few percentage of all mainstream music songs). But these handicaps have a reason to be: the importance and indispensable necessity of constructing a data-set with knowledge objectives. In this case, if timbre evolution of mainstream music has to be studied, it is not acceptable to work with whatever database although if it is very big. And the development and obtainment of the constructed data-set have been long tasks.

Another handicap has been the ground truth obtainment in order to develop the timbre characterization and classification models. That's because labeling perceptually a large amount of songs is a heavy task. But this has to be done if a ground truth of the data-set does not exist.

With all of this an important conclusion can be extracted: future developments with the same will of characterizing timbre evolution could use the same methodology but increasing the data-set in order to improve the results. In the opinion of the author, this is the best option and the first that has to be tested, because at first sight the principal problem is the lack of data, the labeled and the unlabeled. With a bigger data-set the performance of the classifiers could be tested with more warranties. Then, if the performance is bad, other important processes could be reviewed, improved and adjusted: these are the related to characterize and classify timbre. In this sense, future work could use other musical descriptors and other machine learning classification models, such as Gaussian mixture model (GMM) or k-nearest neighbors algorithm (k-NN). And they have to require better accuracy, precision, recall and F-measure measures.

The complete code of this project is published as a contribution to the scientific community. It can be found at https://github.com/aleupons/tfg-timbre-evolution.

Appendix A

Additional information about the work plan

A.1 Work Packages and tasks

Project: Research and investigation	**WP ref:** WP1
Major constituent: Research	Sheet 1 of 5
Short description: This part is the first to be realized to begin the project. It consists in an autonomous learning and research about previous investigations in music timbre field, in order to develop a system able to characterize music timbre. It also consists in a musicological research of the most important moments in popular music which produced important changes or revolutions, especially regarding music timbre. Internal tasks 1 and 2 have to be done before the initialization of next work package.	**Planned start date:** 01/02/2017 **Planned end date:** 31/03/2017 **Start event:** 01/02/2017 **End event:** 23/04/2017
Internal task T1: Investigate of the timbre state of the art. **Internal task T2:** Research the most significant revolutions in popular music history, their genres rise associated and their timbral characteristics, according to musicological bibliography.	**Deliverables:** Project proposal and work plan (28/02/2017) State of the art (23/04/2017)

Table A.1: Work Package 1

Project: Ground truth	**WP ref:** WP2
Major constituent: Database construction and perceptual study	Sheet 2 of 5
Short description: This part consists in the obtainment of ground truth data. This data is the set of lossless audio files that then will be analysed (first perceptually and second analytically), so it has to be representative in terms of timbre characteristics. The database songs are those that achieve first position of Billboard Hot 100 any week of analysed years. Then, these songs have to be listened and perceptually analysed.	**Planned start date:** 27/03/2017 **Planned end date:** 31/03/2017 **Start event:** 23/04/2017 **End event:** 05/05/2017
Internal task T1: Download the Billboard Hot 100's first positions songs of all weeks of the analysed years in order to obtain the database. **Internal task T2:** Choice representative timbral characteristics of each group of songs (each group of years) and classify (binary) perceptually every song.	**Deliverables:** Database: lossless audio files with songs (28/04/2017) Database and perceptual study spreadsheet (05/05/2017)

Table A.2: Work Package 2

Project: System design and implementation	**WP ref:** WP3
Major constituent: Software design, software implementation and programming	Sheet 3 of 5
Short description: This part is the technical one. It consists in the design, implementation and development of the system able to characterize music timbre of songs analysed. It is important to take into account the audio descriptors implemented before and to design the system appropriately according to project objectives and database analysed (those timbre characteristics that we want to describe).	**Planned start date:** 06/05/2017 **Planned end date:** 25/06/2017 **Start event:** 06/05/2017 **End event:** 23/06/2017
Internal task T1: Timbre characterization and feature extraction (development and implementation of the timbre characterization system, in order to extract audio features corresponding to music timbre). **Internal task T2:** Machine learning training, test and classification (using different algorithms and systems) in order to obtain the best possible results according to perceptual analysis. **Internal task T3:** Application of the classification systems to all database songs.	**Deliverables:** Project critical review (07/05/2017) Implemented system: songs features, SVM models, train and test results, classification labels (23/06/2017)

Table A.3: Work Package 3

Project: Evolutive interpretation of the results	**WP ref:** WP4
Major constituent: Interpretation and graphic design	Sheet 4 of 5
Short description: This part is the interpretative one. The objective is to make an interpretation/discussion of the results obtained according to all factors in play (type of music analysed, temporal and historic moments took into account...). It is necessary to trace an evolution of the timbre characteristics of mainstream music.	**Planned start date:** 12/06/2017 **Planned end date:** 30/06/2017 **Start event:** 12/06/2017 **End event:** 30/06/2017
Internal task T1: Graphic design (results presented in a visual way). **Internal task T2:** Discussion and conclusions of the project according to initial assumptions from a musicological/social/cultural point of view.	**Deliverables:** Final report (30/06/2017)

Table A.4: Work Package 4

Project: Documentation	**WP ref:** WP5
Major constituent: Writing	Sheet 5 of 5
Short description: This part has to be developed during all project and consists in the elaboration of the documents related to the thesis. These documents have to guide and plan all tasks, objectives, requirements and specifications at the beginning and during the project.	**Planned start date:** 01/02/2017 **Planned end date:** 30/06/2017 **Start event:** 01/02/2017 **End event:** 30/06/2017
Internal task T1: Write the proposal and work plan. **Internal task T2:** Project critical review redaction. **Internal task T3:** Final report elaboration.	**Deliverables:** Project proposal and work plan (28/02/2017) Project critical review (07/05/2017) Final report (30/06/2017)

Table A.5: Work Package 5

A.2 Milestones

WP	Task	Short title	Milestone/deliverable	Date
1	1	State of the art	State of the art document	03/04/2017
1	2	Significant revolutions in popular music	Modified state of the art document	23/04/2017
2	1	Database songs obtainment	Data-set	28/04/2017
2	2	Perceptual analysis	Perceptual study spreadsheet	05/05/2017
3	1	Timbre characterization	Songs timbre features	10/06/2017
3	2	Machine learning	SVM models, train and test results	23/06/2017
3	3	Classification	Classification results	23/06/2017
4	1	Graphic design	Timbre evolution graphics	27/06/2017
4	2	Discussion and conclusion	Final report	30/06/2017
5	1	Proposal and work plan	Proposal and work plan	28/02/2017
5	2	Project critical review	Project critical review	07/05/2017
5	3	Final report	Final report	30/06/2017

Table A.6: Milestones

Appendix B

Additional information about state of the art

B.1 Timbre study in music information retrieval (MIR)

In [13] and [14] the importance of instrument classification since the beginning of MIR is proved. These studies classified steady-state portions of musical instrument tones in an exemplar-learning-based approach by relying on spectral descriptors such as higher-order moments and amplitudes of spectral peaks. Classification performance was around 50% with more than 60 spectral descriptors and dropped to 42% using only four descriptors (fundamental frequency and the first three spectral moments).

Simultaneously, K.D. Martin modeled perceptual sound source recognition [33], inspired by the hierarchy of perceptual processing proposed by McAdams [36]. The idea was to model source recognition as a process that incrementally accumulates information at multiple, increasingly fine-grained levels of abstraction. A large number of descriptors were computed using a three-dimensional audio representation based on auditory-filterbank autocorrelation by Ellis [10]. The descriptors characterized the spectral, attack, pitch, vibrato and tremolo of the audio. The classification scheme comprised a hierarchical Bayesian decision tree with three levels: all instruments, instrumental families and specific instruments. Classification was realized via a log-likelihood decision that ruled out alternative categories at every level of the tree. With around 75% classification accuracy for instruments, the system performed better than human subjects for specific instruments, whereas humans performed better at the family level.

Other researchers studied which descriptors and classification strategies are more suitable for characterizing timbre. In [24] the authors did it for 0.5 s solo-instruments signals duration with 160 descriptors -including Mel-frequency cepstral coefficients (MFCCs; see section 2.3.1) and octave-band signal intensities and octave-band signal-intensity ratios-, comparing two different feature-selection strategies.

C. Joder et al. also compared different kernel functions for an SVM classifier in [24]: using a radial basis function kernel instead of linear kernel the classification accuracy improved from 81% to 87%. Their last experiment was to increase the signal duration to see its influence on accuracy, so they could realize that the increase from 0.5 to 3 s yielded gain to 93% classification accuracy.

But MFCCs for timbre characterization purpose had already been used before by A. Eronen in [12]. MFCCs are obtained by computing the logarithm of the power of a Mel-scale-warped STFT before applying a discrete cosine transform (DCT). Eronen used the first few MFCC coefficients to capture spectral envelope information.

In [25] the authors considered temporal integration in the descriptor and classification stages, which involves the combination of descriptor observations over successive time frames.

The list of low-level descriptors underwent early and late temporal integration. In first case the new descriptor vectors are computed characterizing the signal at a higher time scale by summing local descriptors extracted from a sequence of analysis frames; while late integration does not extract descriptor dynamics, but either combines successive primary decisions of the classifier or uses a classifier that can deal with sequences. Including both early and late temporal integration yielded small improvements of classification accuracy compared to static reference systems (GMM or SVM with Gaussian kernels) in conjunction with non-integrated features.

B.1.1 MIR, music evolution and musical descriptors efficiency

Research about musical timbre has been also carried out by several authors in different MIR works which also study music evolution or musical descriptors efficiency. So some of them are presented below.

An interesting study about Western popular music evolution along last decades is [45]. In a similar way as the purpose of this work the authors try to characterize and interpret timbre evolution (loudness and pitch are also studied). Timbre characterization -as well as loudness and pitch characterizations- is obtained by encoding data-set descriptions (a collection of audio descriptions and meta-data for a million contemporary music tracks from 1955 to 2010) by a discretization of their values, yielding what they call music *codewords*. To quantify long-term variations of a vocabulary (*codewords*), Monte Carlo sampling in a moving window fashion is performed to obtain vocabulary samples at different periods of time. In particular, for each year, one million beat-consecutive codewords are sampled, considering entire tracks and using a window length of 5 years.

Timbre information of a given frame or beat is originally provided in the million song data-set as an array of 12 real-valued numbers. These numbers correspond to the projection of the (Fourier-based) spectro-temporal representation of the frame's signal into a set of 12 bi-variate basis. These bi-variate basis correspond to "high level abstractions of the spectral surface, ordered by degree of importance". This way, the second basis emphasizes sound brightness, the third is correlated to flatness, the fourth represents sounds with a strong attack, etc.

Authors conclude that Western popular music is becoming more homogeneous as regards harmony aspect, and more restrictive as regards types of timbric patterns it uses. They also note an increasing loudness level of the songs analysed. This research also discovered a cyclic behaviour in last fifty years which shows a predominance of the same timbre features during certain periods of time.

Matthias Mauch et al. made a study very similar to the previous one, which objective was also to characterize timbre and harmony evolution in popular music [34]. However, it not only does not observe the same but flatly conclude that musical diversity -timbral diversity- has not declined in recent decades. This work, which also analyses harmony of the compositions, is focused on the study of American mainstream music (US Billboard Hot 100 between 1960 and 2010) and characterizes its evolution over the last few years, observing big harmonic and timbral changes matching with musical revolutions of certain dates (changes in prevailing music genres).

It uses 14 timbre descriptors to measure the 30-s-long segments of the songs (those appeared in the US Billboard Hot 100 between 1960 and 2010) for a series of quantitative audio features. The timbre features consist of 12 MFCCs, one delta-MFCC value, and one Zero-crossing Count (ZCC) feature. For every frame, they provide a low-dimensional parameterization of the overall shape of the signal's Mel-spectrum, i.e. a spectral representation that takes into account human near-logarithmic perception of sound in magnitude (log-magnitude) and frequency (Mel scale). The first 12 MFCCs (excluding the 0th component) are used and additionally one delta-MFCC, calculated as the difference between any two consecutive values of the 0th MFCC component. These are then discretized into *words* resulting in a timbral lexicon (T-lexicon) of timbre clusters. Each song is represented as a distribution over eight timbral topics (T-topics) that capture particular timbres (e.g. 'drums, aggressive, percussive', 'female voice, melodic, vocal', derived from the expert annotations).

Last example is a very different work without relation on musical timbre evolution. In [40] Timbre Toolbox (a comprehensive set of descriptors that can be useful in perceptual research, as well as in MIR and machine-learning approaches to content-based retrieval in large sound databases) is used to analyse a large database of musical sounds and to assess the informational redundancy of Timbre Toolbox descriptors within the analysed corpus of musical signals based on their intercorrelations. The objective of this article is to quantify the similarity of the various descriptors to estimate approximately the number of groups of statistically independent descriptors, and so assess the extent to which between-descriptor similarities are affected by a change in two important parameters of the analysis pipeline.

The analysis of the independence of audio descriptors is focused on the correlations among descriptors, where pairs of descriptors characterized by a large absolute correlation also share a large amount of information concerning the sound signal. First of all the analysis compares the extent to which a choice of input representation and a choice of descriptive statistic for time-varying descriptors affects the structure of the correlations among the audio descriptors. If various input representations yield highly similar structures for the between-descriptor correlations, they will likely perform similarly in a variety of tasks such as explaining the ratings of participants in a timbre perception study, for example. Secondly, it was created two models that describe the correlational distances among the audio descriptors implemented in the Timbre Toolbox: a hierarchical clustering solution and a multidimensional scaling (MDS) representation. MDS is a means of visualizing the level of similarity of individual cases of a data-set which refers to a set of related ordination techniques used in particular to display the information contained in a distance matrix. The goal of a final analysis was to estimate approximately the number of groups of independent descriptors implemented in the Timbre Toolbox as derived from the cluster analysis.

The authors carried out three different analysis to calculate intercorrelations among descriptors: (1) comparison of the effects of the different input representations and descriptive statistics of the time-varying descriptors on the correlation distances among descriptors; (2) development of distance models for the matrix of between-descriptor correlations; and (3) assessment of the number of groups of independent audio descriptors.

B.2 Timbre study in music psychology

Music psychology assume the qualitative facet of timbre by probing timbral similarity. For this purpose MDS has been important. MDS generates a spatial configuration of points whose pairwise distances approximate the original perceptual dissimilarity data. But there is a variant of this algorithm, CLASCAL, which includes latent classes of subjects weighting the obtained dimensions differently, as well as so-called specificities, which provide additional distance values to account for perceptual features that are specific to individual items, taking up as much unexplained variance as possible but not making any assumptions about the relationships among timbres.

The application of MDS to timbre similarity perception -first conducted by Plomp [43]- needs the avoid of interferences from other perceptual parameters. In this way stimuli are subjectively equalized in pitch, loudness, and duration before participants are asked to judge dissimilarity of subsequently presented pairs of timbres, corresponding to the researcher the search for physical correlates of the obtained spatial dimensions. Commonly these experiments use small set of timbres (10-20) due to the large number of pairs to be compared. But recently 42 tones were used by allocating different sparse subgroups of sounds [9]. The full dissimilarity matrix was obtained by averaging across groups.

One example of an MDS application is the one conducted by John Grey in his seminal work on timbre [17, 18, 19], where he used emulations of orchestral tones generated by line-segment-approximated amplitude and frequency trajectories of partials. He settled on a three-dimensional MDS solution. The physical correlates were interpreted in terms of properties of the spectral energy distribution for the dimension one. Dimension two was related to the attack synchronicity of partials, but he also noticed that this dimension was related to the amount of spectral fluctuation. Dimension three was related to spectral balance during the attack portion of tones.

A few years later Krumhansl was the first to present a timbre space including specificities using CLASCAL [28]. MDS dimension one was interpreted qualitatively as corresponding to rapidity of attack, dimension two to centre of gravity of the spectrum and dimension three to spectral fluctuations over time. Then, recorded instrumental sounds were used and the influence of attack portions on similarity judgments were studied [23]. For all three sets of stimuli, similarity judgments correlated with spectral centroid frequency -first moment of the spectral distribution and the centre of gravity of the spectrum correlates with subjective brightness- and amplitude envelope shape.

Another MDS application combining CLASCAL is [37], where authors synthesized many of the possibilities of MDS, including specificities plus latent classes of subjects using CLASCAL, as well as rigorous quantification of physical correlates of MDS dimensions. The audio descriptors log-rise time (logarithm of duration from start of tone to amplitude maximum), spectral centroid, spectral flux (average of correlations between adjacent short-time amplitude spectra), and spectral irregularity (log of the standard deviation of component amplitudes of a tone's spectral envelope, derived from a running average of the amplitudes of three adjacent harmonics) were considered for an interpretation of a CLASCAL-based MDS model. However, the best model fit was obtained by a six-dimensional solution without specificities that yielded an ambiguous acoustic interpretation. The authors thus settled on a three-dimensional solution that was easier to interpret psychophysically. Here, dimension one and rise time and dimension two and spectral centroid both had correlations of

0.94. Dimension three had a correlation of 0.54 with spectral flux. This study confirmed the salience of the spectral centroid and amplitude envelope properties, but it also highlighted the interpretative role of the researcher using MDS.

In Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones (2005), Caclin, McAdams, Smith and Winsberg addressed the issue of correlation versus causation in timbre-space studies. They synthesized tones varying along the spectral centroid, rise time and spectral flux parameters. Spectral flux was operationalized as variation of spectral centroid within the first 100 ms of the tone. Spectral deviation was confirmed to be perceptually salient in another experiment in that paper, but the obtained timbre spaces suggested that spectral flux is unlikely to serve as a salient perceptual dimension of timbre, at least in the parameterization used for the experiment.

Finally, there is a different approximation to timbre similarity and classification which discarded MDS altogether and modeled dissimilarity judgements directly: [27]. Dissimilarities were modelled by fitting a Gaussian kernel distance on a high-dimensional time x frequency x modulation-rate x modulation-scale representation based on estimates of spectrotemporal receptive fields (STRF) modelled on auditory cortical neurons. After a complex kernel optimization was performed they obtained correlations of r = 0.94 with human perceptual judgements. Using simple Euclidean norms instead, correlation with dissimilarity judgements reduced to r = 0.61. At the same time, instrument classification accuracies using SVM were above 95%. This demonstrates that there is enough information in STRFs to accomplish the tasks of timbre classification and dissimilarity prediction.

Appendix C

Constructed data-set and its perceptual classification

1962-1966: British pop		
Songs	**Weeks on Hot 100's 1st position**	**Vocal accompaniment (choruses)**
1962		
Bobby Boris Pickett And The Crypt-Kickers - Monster Mash	2	yes
Bobby Vinton - Roses Are Red (My Love)	4	yes
Bruce Channel - Hey! Baby	3	no
Chubby Checker - The Twist	2	yes
Connie Francis - Don't Break The Heart That Loves You	1	no
David Rose & His Orchestra - The Stripper	1	no
Elvis Presley With The Jordanaires - Good Luck Charm	2	yes
Gene Chandler - Duke Of Earl	3	yes
Joey Dee & The Starliters - Peppermint Twist Part I	3	no
Little Eva - The Loco-Motion	1	yes
Mr. Acker Bilk - Stranger On The Shore	1	no
Neil Sedaka - Breaking Up Is Hard To Do	2	yes
Ray Charles - I Can't Stop Loving You	5	no
Shelley Fabares - Johnny Angel	2	yes
The 4 Seasons - Big Girls Don't Cry	5	yes
The 4 Seasons - Sherry	5	yes
The Crystals - He's A Rebel	2	no
The Shirelles - Soldier Boy	3	yes
The Tokens - The Lion Sleeps Tonight	1	yes
The Tornados - Telstar	2	no
Tommy Roe - Sheila	2	yes
1963		
Bobby Vinton - Blue Velvet	3	yes
Dale & Grace - I'm Leaving It Up To You	2	no
Jan & Dean - Surf City	2	yes
Jimmy Gilmer And The Fireballs - Sugar Shack	5	no
Jimmy Soul - If You Wanna Be Happy	2	no
Kyu Sakamoto - Sukiyaki	3	no
Lesley Gore - It's My Party	2	yes
Little Peggy March - I Will Follow Him	3	yes
Little Stevie Wonder - Fingertips - Pt 2	3	no
Nino Tempo & April Stevens - Deep Purple	1	yes
Paul and Paula - Hey Paula	3	no
Ruby & The Romantics - Our Day Will Come	1	yes
Steve Lawrence - Go Away Little Girl	2	no

The 4 Seasons - Walk Like A Man	3	yes
The Angels - My Boyfriend's Back	3	yes
The Chiffons - He's So Fine	4	yes
The Essex - Easier Said Than Done	2	yes
The Rooftop Singers - Walk Right In	2	no
The Singing Nun (Soeur Sourire) - Dominique	4	no
The Tornados - Telstar	1	no
The Tymes - So Much In Love	1	yes
1964		
Bobby Vinton - Mr. Lonely	1	yes
Bobby Vinton - There! I've Said It Again	4	yes
Dean Martin - Everybody Loves Somebody	1	no
Lorne Greene - Ringo	1	yes
Louis Armstrong And The All Stars - Hello, Dolly!	1	no
Manfred Mann - Do Wah Diddy Diddy	2	no
Mary Wells - My Guy	2	yes
Peter And Gordon - A World Without Love	1	no
Roy Orbison And The Candy Men - Oh, Pretty Woman	3	no
The 4 Seasons Featuring The Sound Of Frankie Valli - Rag Doll	2	yes
The Animals - The House of the Rising Sun	3	no
The Beach Boys - I Get Around	2	yes
The Beatles - A Hard Day's Night	2	yes
The Beatles - Can't Buy Me Love	5	no
The Beatles - I Feel Fine	1	yes
The Beatles - I Want To Hold Your Hand	7	no
The Beatles - Love Me Do	1	no
The Beatles - She Loves You	2	no
The Dixie Cups - Chapel Of Love	3	no
The Shangri-Las - Leader Of The Pack	1	yes
The Supremes - Baby Love	4	yes
The Supremes - Come See About Me	1	yes
The Supremes - Where Did Our Love Go	2	yes
1965		
Barry McGuire - Eve Of Destruction	1	no
Four Tops - I Can't Help Myself (Sugar Pie Honey Bunch)	2	yes
Freddie And The Dreamers - I'm Telling You Now	2	yes
Gary Lewis And The Playboys - This Diamond Ring	2	no
Herman's Hermits - I'm Henry VIII, I Am	1	no
Herman's Hermits - Mrs. Brown You've Got A Lovely Daughter	3	yes
Petula Clark - Downtown	2	no
Sonny & Cher - I Got You Babe	3	no
The Beach Boys - Help Me, Rhonda	2	yes
The Beatles - Eight Days a Week	2	no
The Beatles - Help!	3	yes
The Beatles - I Feel Fine	2	yes
The Beatles - Ticket to Ride	1	no
The Beatles - Yesterday	4	no
The Byrds - Mr. Tambourine Man	1	yes
The Byrds - Turn! Turn! Turn! (To Everything There Is A Season)	3	no

The Dave Clark Five - Over And Over	1	no
The McCoys - Hang On Sloopy	1	yes
The Righteous Brothers - You've Lost That Lovin' Feelin'	2	no
The Rolling Stones - (I Can't Get No) Satisfaction	4	no
The Rolling Stones - Get Off of My Cloud	2	no
The Supremes - Back In My Arms Again	1	yes
The Supremes - Come See About Me	1	yes
The Supremes - I Hear A Symphony	2	yes
The Supremes - Stop! In The Name Of Love	2	no
The Temptations - My Girl	1	no
Wayne Fontana And The Mindbenders - Game Of Love	1	yes
1966		
Donovan - Sunshine Superman	1	no
Four Tops - Reach Out I'll Be There	2	yes
Frank Sinatra - Stranges In The Night	1	no
Johnny Rivers - Poor Side Of Town	1	yes
Lou Christie - Lightnin' Strikes	1	no
Nancy Sinatra - These Boots Are Made For Walkin'	1	no
Percy Sledge - When A Man Loves A Woman	2	no
Petula Clark - My Love	2	no
Question Mark & The Mysterians - 96 Tears	1	no
Simon & Garfunkel - The Sound Of Silence	2	no
Ssgt Barry Sadler - The Ballad Of The Green Berets	5	yes
The Association - Cherish	3	no
The Beach Boys - Good Vibrations	1	yes
The Beatles - Paperback Writer	2	yes
The Beatles - We Can Work It Out	3	no
The Lovin' Spoonful - Summer In The City	3	no
The Mamas & The Papas - Monday, Monday	3	yes
The Monkees - I'm A Believer	1	yes
The Monkees - Last Train To Clarksville	1	yes
The New Vaudeville Band - Winchester Cathedral	3	no
The Righteous Brothers - (You're My) Soul And Inspiration	3	yes
The Rolling Stones - Paint It, Black	2	no
The Supremes - You Can't Hurry Love	2	yes
The Supremes - You Keep Me Hangin' On	2	yes
The Troggs - Wild Thing	2	no
The Young Rascals - Good Lovin'	1	yes
Tommy James And The Shondells - Hanky Panky	2	no

Table C.1: 1962-1966 data-set and its perceptual classification

1975-1979: Disco & Punk				
Songs	**Weeks on Hot 100's 1st position**	**Backing keyboard instruments**	**Orchestral instruments**	**Distorted power chords**
1975				
America - Sister Golden Hair	1	no	no	no

AWB - Pick Up The Pieces	1	yes	yes	no
B.J. Thomas - (Hey Won't You Play) Another Somebody Done Somebody Wrong Song	1	yes	no	no
Barry Manilow - Mandy	1	yes	yes	no
Bee Gees - Jive Talkin'	2	yes	no	no
Captain & Tennille - Love Will Keep Us Together	4	yes	no	no
Carpenters - Please Mr. Postman	1	yes	no	no
David Bowie - Fame	2	no	no	yes
Eagles - Best of My Love	1	no	no	no
Eagles - One Of These Nights	1	no	no	no
Earth, Wind & Fire - Shining Star	1	yes	yes	no
Elton John - Island Girl	3	yes	no	no
Elton John - Lucy In The Sky With Diamonds	2	yes	no	no
Frankie Valli - My Eyes Adored You	1	yes	no	no
Freddy Fender - Before The Next Teardrop Falls	1	no	no	no
Glen Campbell - Rhinestone Cowboy	2	yes	yes	no
Hamilton, Joe Frank & Reynolds - Fallin' In Love	1	yes	yes	no
John Denver - I'm Sorry	1	no	yes	no
John Denver - Thank God I'm A Country Boy	1	no	no	no
KC And The Sunshine Band - Get Down Tonight	1	yes	yes	no
KC And The Sunshine Band - That's The Way (I Like It)	2	yes	yes	no
Labelle - Lady Marmalade	1	yes	yes	no
Linda Ronstadt - You're No Good	1	yes	no	no
Minnie Riperton - Lovin' You	1	no	yes	no
Neil Sedaka - Bad Blood	3	no	yes	no
Neil Sedaka - Laughter In The Rain	1	yes	yes	no
Ohio Players - Fire	1	no	yes	no
Olivia Newton-John - Have You Never Been Mellow	1	no	yes	no
Silver Convention - Fly Robin, Fly	3	yes	yes	no
The Doobie Brothers - Black Water	1	no	no	no
The Elton John Band - Philadelphia Freedom	2	yes	yes	no
The Staple Singers - Let's Do It Again	1	no	yes	no
Tony Orlando & Dawn - He Don't Love You Like I Love You	3	no	yes	no
Van McCoy And The Soul City Symphony - The Hustle	1	yes	yes	no
Wings - Listen To What The Man Said	1	no	no	no
1976				
Barry Manilow - I Write The Songs	1	yes	yes	no
Bay City Rollers - Saturday Night	1	yes	no	no
Bee Gees - You Should Be Dancing	1	yes	yes	no
Bellamy Brothers - Let Your Love Flow	1	no	no	no
Chicago - If You Leave Me Now	2	no	yes	no
CW McCall - Convoy	1	no	yes	no
Diana Ross - Love Hangover	2	yes	yes	no
Diana Ross - Theme From Mahogany (Do You Know Where You're Going To)	1	yes	yes	no

Elton John & Kiki Dee - Don't Go Breaking My Heart	4	yes	yes	no
John Sebastian - Welcome Back	1	yes	no	no
Johnnie Taylor - Disco Lady	4	yes	yes	no
KC And The Sunshine Band - (Shake, Shake, Shake) Shake Your Booty	1	yes	yes	no
Ohio Players - Love Rollercoaster	1	yes	yes	no
Paul Simon - 50 Ways To Leave Your Lover	3	no	no	no
Rhythm Heritage - Theme From S.W.A.T.	1	yes	yes	no
Rick Dees & His Cast Of Idiots - Disco Duck (Part I)	1	no	yes	no
Rod Stewart - Tonight's The Night (Gonna Be Alright)	7	no	yes	no
Starland Vocal Band - Afternoon Delight	2	no	yes	no
Steve Miller - Rock'n Me	1	no	no	no
The 4 Seasons - December, 1963 (Oh, What A Night)	3	yes	yes	no
The Manhattans - Kiss And Say Goodbye	2	no	yes	no
The Miracles - Love Machine [Part 1]	1	yes	yes	no
The Sylvers - Boogie Fever	1	yes	yes	no
Walter Murphy & The Big Apple Band - A Fifth Of Beethoven	1	yes	yes	no
Wild Cherry - Play That Funky Music	3	yes	yes	no
Wings - Silly Love Songs	5	yes	yes	no
1977				
ABBA - Dancing Queen	1	yes	yes	no
Alan O'Day - Undercover Angel	1	yes	no	no
Andy Gibb - I Just Want To Be Your Everything	4	yes	yes	no
Barbra Streisand - Evergreen (Love Theme From A Star Is Born)	3	no	yes	no
Barry Manilow - Looks Like We Made It	1	yes	yes	no
Bee Gees - How Deep Is Your Love	2	yes	yes	no
Bill Conti - Gonna Fly Now	1	no	yes	no
Daryl Hall And John Oates - Rich Girl	2	no	yes	no
David Soul - Don't Give Up On Us	1	yes	yes	no
Debby Boone - You Light Up My Life	10	yes	yes	no
Eagles - Hotel California	1	yes	no	yes
Eagles - New Kid In Town	1	no	no	no
Fleetwood Mac - Dreams	1	yes	no	no
Glen Campbell - Southern Nights	1	yes	yes	no
KC And The Sunshine Band - I'm Your Boogie Man	1	yes	yes	no
Leo Sayer - When I Need You	1	no	yes	no
Leo Sayer - You Make Me Feel Like Dancing	1	yes	yes	no
Manfred Mann's Earth Band - Blinded By The Light	1	yes	no	no
Marilyn McCoo & Billy Davis Jr. - You Don't Have To Be A Star (To Be In My Show)	1	yes	yes	no
Marvin Gaye - Got To Give It Up (Pt. I)	1	yes	no	no
Mary MacGregor - Torn Between Two Lovers	2	yes	yes	no

Meco - Star Wars Theme-Cantina Band	2	yes	yes	no
Rod Stewart - Tonight's The Night (Gonna Be Alright)	1	no	yes	no
Rose Royce - Car Wash	1	yes	yes	no
Shaun Cassidy - Da Doo Ron Ron	1	yes	yes	no
Stevie Wonder - I Wish	1	no	yes	no
Stevie Wonder - Sir Duke	3	yes	yes	no
The Emotions - Best Of My Love	5	yes	yes	no
Thelma Houston - Don't Leave Me This Way	1	yes	yes	no
1978				
A Taste Of Honey - Boogie Oogie Oogie	3	yes	no	no
Andy Gibb - (Love Is) Thicker Than Water	2	no	no	no
Andy Gibb - Shadow Dancing	7	yes	no	no
Anne Murray - You Needed Me	1	no	yes	no
Barbra Streisand & Neil Diamond - You Don't Bring Me Flowers	2	yes	yes	no
Bee Gees - How Deep Is Your Love	1	yes	yes	no
Bee Gees - Night Fever	8	yes	yes	no
Bee Gees - Stayin' Alive	4	yes	yes	no
Chic - Le Freak	3	yes	yes	no
Commodores - Three Times A Lady	2	yes	yes	no
Donna Summer - MacArthur Park	3	yes	yes	no
Exile - Kiss You All Over	4	yes	yes	no
Frankie Valli - Grease	2	yes	yes	no
John Travolta & Olivia Newton-John - You're The One That I Want	1	yes	yes	no
Johnny Mathis And Deniece Williams - Too Much, Too Little, Too Late	1	no	yes	no
Nick Gilder - Hot Child In The City	1	no	no	no
Player - Baby Come Back	3	yes	no	no
The Rolling Stones - Miss You	1	no	no	yes
Wings - With A Little Luck	2	yes	no	no
Yvonne Elliman - If I Can't Have You	1	no	yes	no
1979				
Amii Stewart - Knock On Wood	1	yes	yes	yes
Anita Ward - Ring My Bell	2	no	no	no
Barbra Streisand & Donna Summers - No More Tears (Enough Is Enough)	2	no	yes	no
Bee Gees - Love You Inside Out	1	yes	yes	no
Bee Gees - Too Much Heaven	2	no	yes	no
Bee Gees - Tragedy	2	yes	yes	no
Blondie - Heart Of Glass	1	yes	no	no
Chic - Good Times	1	yes	yes	no
Chic - Le Freak	3	yes	yes	no
Commodores - Still	1	yes	yes	no
Donna Summer - Bad Girls	5	yes	yes	no
Donna Summer - Hot Stuff	3	yes	yes	no
Doobie Brothers - What A Fool Believes	1	yes	no	no
Eagles - Heartache Tonight	1	no	no	yes
Gloria Gaynor - I Will Survive	3	yes	yes	no

Herb Alpert - Rise	2	yes	yes	no
M - Pop Muzik	1	yes	yes	no
Michael Jackson - Don't Stop 'Til You Get Enough	1	yes	yes	no
Peaches & Herb - Reunited	4	no	yes	no
Robert John - Sad Eyes	1	yes	yes	no
Rod Stewart - Da Ya Think I'm Sexy	4	yes	yes	no
Rupert Holmes - Escape (The Pina Colada Song)	2	yes	no	no
Styx - Babe	2	yes	yes	no
The Knack - My Sharona	6	no	no	yes

Table C.2: 1975-1979 data-set and its perceptual classification

1989-1993: Hip hop			
Songs	**Weeks on Hot 100's 1st position**	**Drum machine**	**High load of low frequencies**
1989			
Bad English - When I See You Smile	2	no	no
Bangles - Eternal Flame	1	no	no
Bette Midler - Wind Beneath My Wings (From 'Beaches')	1	no	no
Billy Joel - We Didn't Start The Fire	2	no	yes
Bobby Brown - My Prerogative	1	yes	yes
Bon Jovi - I'll Be There For You	1	no	no
Debbie Gibson - Lost In Your Eyes	3	no	no
Fine Young Cannibals - Good Thing	1	no	yes
Fine Young Cannibals - She Drives Me Crazy	1	yes	yes
Gloria Estefan - Don't Wanna Lose You Now	1	no	no
Janet Jackson - Miss You Much	4	yes	yes
Madonna - Like A Prayer	3	no	no
Martika - Toy Soldiers	2	no	no
Michael Damian - Rock On ('From A Little Dream')	1	no	no
Mike - The Living Years	1	no	no
Milli Vanilli - Baby Don't Forget My Number	1	yes	yes
Milli Vanilli - Blame It on the Rain	2	yes	yes
Milli Vanilli - Girl I'm Gonna Miss You	2	yes	yes
New Kids On The Block - Hangin Tough	1	no	yes
New Kids On The Block - I'll Be Loving You (Forever)	1	no	no
Paula Abdul - Cold Hearted	1	yes	yes
Paula Abdul - Forever Your Girl	2	yes	yes
Paula Abdul - Straight Up	3	yes	yes
Phil Collins - Another Day In Paradise	2	no	no
Phil Collins - Two Hearts	2	no	no
Poison - Every Rose Has Its Thorn	1	no	no
Prince - Batdance (From 'Batman')	1	no	yes
Richard Marx - Right Here Waiting	3	no	no
Richard Marx - Satisfied	1	no	no
Roxette - Listen To Your Heart	1	no	no

Roxette - The Look	1	no	no
Sheriff - When I'm With You	1	no	no
Simply Red - If You Don't Know Me By Now	1	no	no
1990			
Alannah Myles - Black Velvet	2	no	yes
George Michael - Praying For Time	1	no	no
Glenn Medeiros - She Ain't Worth It	2	yes	yes
James Ingram - I Don't Have The Heart	1	no	no
Janet Jackson - Black Cat	1	no	no
Janet Jackson - Escapade	3	yes	yes
Jon Bon Jovi - Blaze Of Glory	1	no	no
Madonna - Vogue	3	yes	no
Mariah Carey - Love Takes Time	3	no	no
Mariah Carey - Vision Of Love	4	no	no
Maxi Priest - Close To You	1	no	yes
Michael Bolton - How Am I Supposed To Live Without You	3	no	no
Nelson - (Can't Live Without Your) Love And Affection	1	no	no
New Kids On The Block - Step By Step	3	yes	no
Paula Abdul - Opposites Attract	3	yes	yes
Phil Collins - Another Day In Paradise	2	no	no
Roxette - It Must Have Been Love	2	no	no
Sinead O Connor - Nothing Compares 2 U	4	no	no
Stevie B - Because I Love You	4	no	no
Sweet Sensation - If Wishes Came True	1	no	no
Taylor Dayne - Love Will Lead You Back	1	no	no
Tommy Page - I'll Be Your Everything	1	no	no
Vanilla Ice - Ice Ice Baby	1	yes	yes
Whitney Houston - I'm Your Baby Tonight	1	yes	no
Wilson Phillips - Hold On	1	no	no
Wilson Phillips - Release Me	2	no	no
1991			
Amy Grant - Baby Baby	2	yes	no
Bryan Adams - (Everything I Do) I Do It For You (From 'Robin Hood')	7	no	no
C+C Music Factory - Gonna Make You Sweat	2	yes	yes
Color Me Badd - I Adore Mi Amor	2	yes	no
EMF - Unbelievable	1	no	yes
Extreme - More Than Words	1	no	no
Gloria Estefan - Coming Out Of The Dark	2	no	no
Hi-Five - I Like The Way (The Kissing Game)	1	yes	no
Janet Jackson - Love Will Never Do (Without You)	1	yes	yes
Karyn White - Romantic	1	yes	no
Londonbeat - I've Been Thinking About You	1	yes	no
Madonna - Justify My Love	2	yes	yes
Mariah Carey - Emotions	3	no	no
Mariah Carey - I Don't Wanna Cry	2	no	no
Mariah Carey - Someday	2	yes	no
Marky Mark - Good Vibrations	1	yes	yes
Michael Bolton - When A Man Loves A Woman	1	no	no

Michael Jackson - Black Or White	4	yes	no
P.M. Dawn - Set Adrift On Memory Bliss	1	yes	no
Paula Abdul - Rush Rush	5	yes	no
Paula Abdul - The Promise Of A New Day	1	yes	no
Prince And The N.P.G. - Cream	2	no	no
Roxette - Joyride	1	no	no
Surface - The First Time	2	no	no
Timmy T. - One More Try	1	yes	no
Whitney Houston - All The Man That I Need	2	no	no
Wilson Phillips - You're In Love	1	no	no
1992			
Boyz II Men - End Of The Road	13	no	no
Color Me Badd - All 4 Love	1	no	no
George Michael & Elton John - Don't Let The Sun Go Down On Me	1	no	no
Kris Kross - Jump	8	yes	yes
Madonna - This Used To Be My Playground	1	yes	no
Mariah Carey - I'll Be There	2	no	no
Michael Jackson - Black Or White	3	yes	no
Mr. Big - To Be With You	3	no	no
Right Said Fred - I'm Too Sexy	3	yes	yes
Sir Mix-A-Lot - Baby Got Back	5	yes	yes
The Heights - How Do You Talk To An Angel	2	no	no
Vanessa William - Save The Best For Last	5	no	no
Whitney Houston - I Will Always Love You (From 'The Bodyguard')	5	no	no
1993			
Janet Jackson - Again	2	no	no
Janet Jackson - That's The Way Love Goes	8	yes	yes
Mariah Carey - Dreamlover	8	yes	yes
Mariah Carey - Hero	1	no	no
Meat Loaf - I'd Do Anything For Love (But I Won't Do That)	5	no	no
Peabo Bryson - A Whole New World (Aladdin's Theme)	1	no	no
Silk - Freak Me	2	yes	yes
Snow - Informer	7	yes	no
SWV - Weak	2	no	no
UB40 - Can't Help Falling In Love	7	yes	yes
Whitney Houston - I Will Always Love You (From 'The Bodyguard')	9	no	no

Table C.3: 1989-1993 data-set and its perceptual classification

1997-2001: Boy/girl bands & Pop starlets			
Songs	**Weeks on Hot 100's 1st position**	**Electronic backing**	**Vocal accompaniment (choruses)**
1997			
Boyz II Men - 4 Seasons Of Loneliness	1	yes	yes

Elton John - Candle In The Wind 1997-Something About The Way You Look Tonight	12	no	no
Hanson - Mmmbop	3	no	yes
Mariah Carey - Honey	3	yes	yes
Puff Daddy Featuring Mase - Can't Nobody Hold Me Down	6	yes	no
Puff Daddy & Faith Evans Featuring 112 - I'll Be Missing You	11	no	no
Spice Girls - Wannabe	4	yes	yes
The Notorious B.I.G. - Hypnotize	3	yes	no
The Notorious B.I.G. Featuring Puff Daddy & Mase - Mo Money Mo Problems	2	yes	no
Toni Braxton - Un-Break My Heart	7	no	no
1998			
Aerosmith - I Don't Want To Miss A Thing	4	no	no
Barenaked Ladies - One Week	1	no	no
Brandy & Monica - The Boy Is Mine	13	yes	yes
Celine Dion - My Heart Will Go On	2	no	no
Divine - Lately	1	yes	yes
Elton John - Candle In The Wind 1997-Something About The Way You Look Tonight	2	no	no
Janet Jackson - Together Again	2	yes	yes
K-Ci & JoJo - All My Life	3	yes	yes
Lauryn Hill - Doo Wop (That Thing)	2	yes	yes
Mariah Carey - My All	1	no	no
Monica - The First Night	5	yes	yes
Next - Too Close	5	yes	yes
R. Kelly - I'm Your Angel	4	yes	no
Savage Garden - Truly Madly Deeply	2	yes	yes
Usher - Nice & Slow	2	yes	yes
Will Smith - Gettin' Jiggy Wit It	3	yes	no
1999			
Brandy - Have You Ever	2	no	yes
Britney Spears - ...Baby One more Time	2	yes	yes
Cher - Believe	4	yes	no
Christina Aguilera - Genie In A Bottle	5	yes	yes
Destiny's Child - Bills, Bills, Bills	1	yes	yes
Enrique Iglesias - Bailamos	2	no	yes
Jennifer Lopez - If You Had My Love	5	yes	no
Mariah Carey Featuring Jay-Z - Heartbreaker	2	yes	yes
Monica - Angel Of Mine	4	yes	yes
R. Kelly - I'm Your Angel	2	yes	no
Ricky Martin - Livin' La Vida Loca	5	yes	no
Santana Featuring Rob Thomas - Smooth	10	no	no
TLC - No Scrubs	4	yes	no
TLC - Unpretty	3	yes	yes
Will Smith Featuring Dru Hill - Wild Wild West	1	no	no
2000			
Aaliyah - Try Again	1	yes	no

Christina Aguilera - Come On Over Baby (All I Want Is You)	4	yes	yes
Christina Aguilera - What A Girl Wants	2	yes	yes
Creed - With Arms Wide Open	1	no	no
Destiny's Child - Independent Women Part 1	7	yes	yes
Destiny's Child - Say My Name	3	yes	yes
Enrique Iglesias - Be With You	3	yes	yes
Janet Jackson - Doesn't Really Matter	3	yes	yes
Lonestar - Amazed	2	no	yes
Madonna - Music	4	yes	no
Mariah Carey Featuring Joe - Thank God I Found You	1	yes	yes
Matchbox Twenty - Bent	1	no	no
N Sync - It's Gonna Be Me	2	yes	yes
Santana Featuring Rob Thomas - Smooth	2	no	no
Santana Featuring The Product G&B - Maria Maria	10	no	no
Savage Garden - I Knew I Loved You	4	no	no
Sisqo - Incomplete	2	yes	no
Vertical Horizon - Everything You Want	1	no	no
2001			
Alicia Keys - Fallin'	6	yes	yes
Christina Aguilera, Lil' Kim, Mya & Pink - Lady Marmalade	5	yes	yes
Crazy Town - Butterfly	2	no	no
Destiny's Child - Bootylicious	2	yes	yes
Destiny's Child - Independent Women Part 1	4	yes	yes
Janet Jackson - All For You	7	yes	yes
Jennifer Lopez Featuring Ja Rule - I'm Real	5	yes	no
Joe Featuring Mystikal - Stutter	4	no	yes
Mary J. Blige - Family Affair	6	yes	yes
Nickelback - How You Remind Me	2	no	no
Outkast - Ms. Jackson	1	yes	yes
Shaggy Featuring Rayvon - Angel	1	yes	no
Shaggy Featuring Ricardo 'RikRok' Ducent - It Wasn't Me	2	yes	no
Usher - U Got It Bad	1	yes	yes
Usher - U Remind Me	4	yes	yes

Table C.4: 1997-2001 data-set and its perceptual classification

2012-2016: Trap			
Songs	**Weeks on Hot 100's 1st position**	**Sub-bass double/triple kick drums**	**Double-time, twitchy hi-hats**
2012			
Adele - Set Fire To The Rain	2	no	no
Bruno Mars - Locked Out Of Heaven	2	no	no
Carly Rae Jepsen - Call Me Maybe	9	no	no
Flo Rida - Whistle	2	no	no
Fun Featuring Janelle Monae - We Are Young	6	no	no

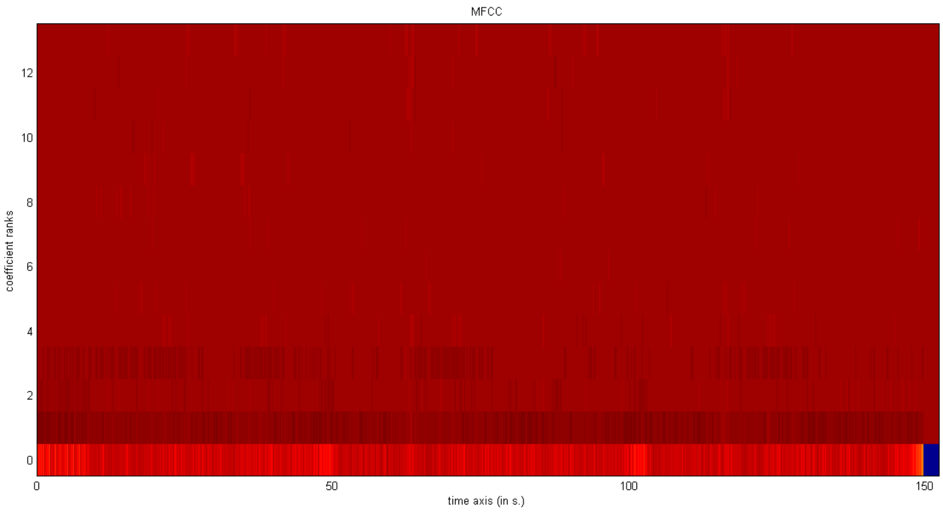
Gotye Featuring Kimbra - Somebody That I Used To Know	8	no	no
Katy Perry - Part Of Me	1	no	no
Kelly Clarkson - Stronger (What Doesn't Kill You)	3	no	no
LMFAO - Sexy And I Know It	2	no	no
Maroon 5 - One More Night	9	yes	no
Rihanna - Diamonds	3	no	yes
Rihanna Featuring Calvin Harris - We Found Love	2	no	no
Taylor Swift - We Are Never Ever Getting Back Together	3	no	yes
2013			
Baauer - Harlem Shake	5	yes	yes
Bruno Mars - Locked Out Of Heaven	4	no	no
Bruno Mars - When I Was Your Man	1	no	no
Eminem Featuring Rihanna - The Monster	2	no	no
Katy Perry - Roar	2	no	no
Lorde - Royals	9	yes	yes
Macklemore & Ryan Lewis Featuring Ray Dalton - Can't Hold Us	5	no	no
Macklemore & Ryan Lewis Featuring Wanz - Thrift Shop	6	yes	no
Miley Cyrus - Wrecking Ball	3	no	no
Pink Featuring Nate Ruess - Just Give Me A Reason	3	no	no
Robin Thicke Featuring T.I. + Pharrell - Blurred Lines	12	no	yes
2014			
Eminem Featuring Rihanna - The Monster	2	no	no
Iggy Azalea Featuring Charli XCX - Fancy	7	yes	yes
John Legend - All Of Me	3	no	no
Katy Perry Featuring Juicy J - Dark Horse	4	yes	no
MAGIC! - Rude	6	no	no
Meghan Trainor - All About That Bass	8	no	no
Pharrell Williams - Happy	10	no	no
Pitbull Featuring Ke$ha - Timber	3	no	no
Taylor Swift - Blank Space	5	yes	no
Taylor Swift - Shake It Off	4	yes	no
2015			
Adele - Hello	7	no	no
Justin Bieber - What Do You Mean	1	yes	no
Mark Ronson Featuring Bruno Mars - Uptown Funk!	14	no	no
OMI - Cheerleader	6	yes	no
Taylor Swift - Blank Space	2	yes	no
Taylor Swift Featuring Kendrick Lamar - Bad Blood	1	no	no
The Weeknd - Can't Feel My Face	3	no	no
The Weeknd - The Hills	6	yes	yes
Wiz Khalifa Featuring Charlie Puth - See You Again	12	yes	no
2016			
Adele - Hello	3	no	no
Desiigner - Panda	2	yes	yes
Drake Featuring WizKid & Kyla - One Dance	10	no	no
Justin Bieber - Love Yourself	2	no	no

Justin Bieber - Sorry	3	yes	yes
Justin Timberlake - Can't Stop The Feeling!	1	no	no
Rae Sremmurd Featuring Gucci Mane - Black Beatles	6	yes	yes
Rihanna Featuring Drake - Work	9	yes	yes
Sia Featuring Sean Paul - Cheap Thrills	4	no	no
The Chainsmokers Featuring Halsey - Closer	12	no	yes
Zayn - Pillowtalk	1	no	yes

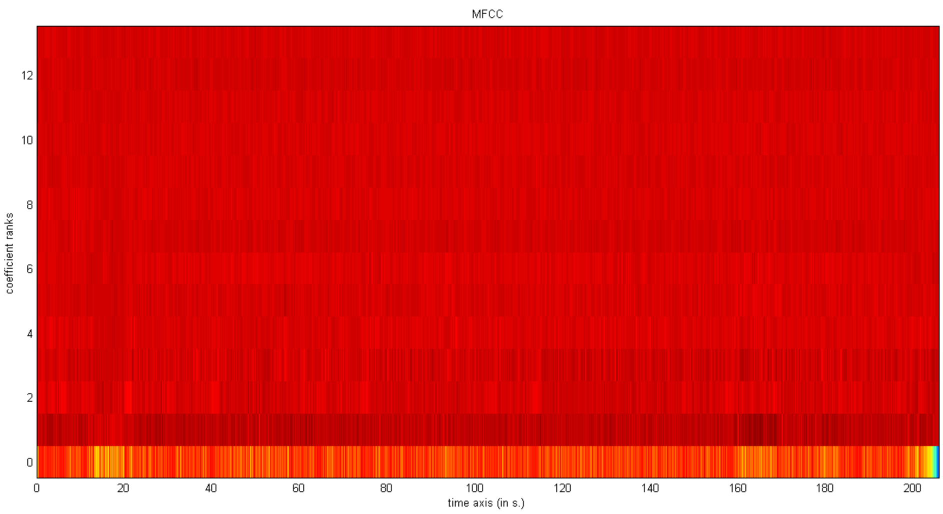
Table C.5: 2012-2016 data-set and its perceptual classification

Appendix D

Timbre characterization results: figures of two examples

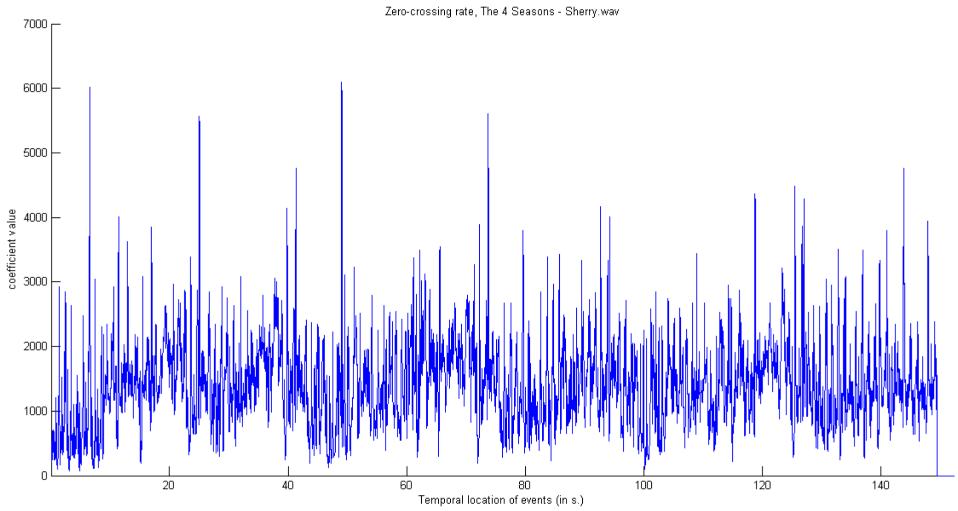


(a) MFCCs computation of *The 4 Seasons - Sherry* (1962)

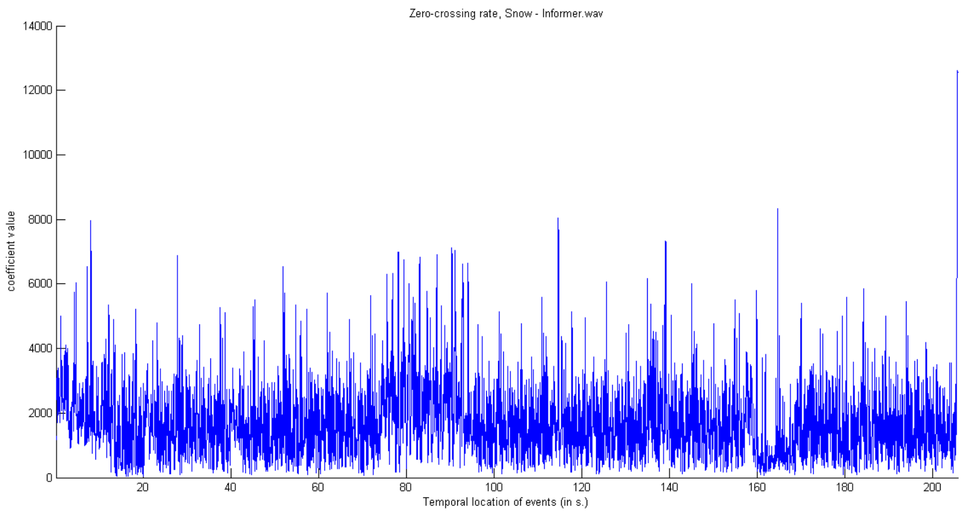


(b) MFCCs computation of *Snow - Informer* (1993)

Figure D.1: Comparison of MFCCs for two different songs. The first 14 MFCCs are computed for frames of 50ms and half overlapping.

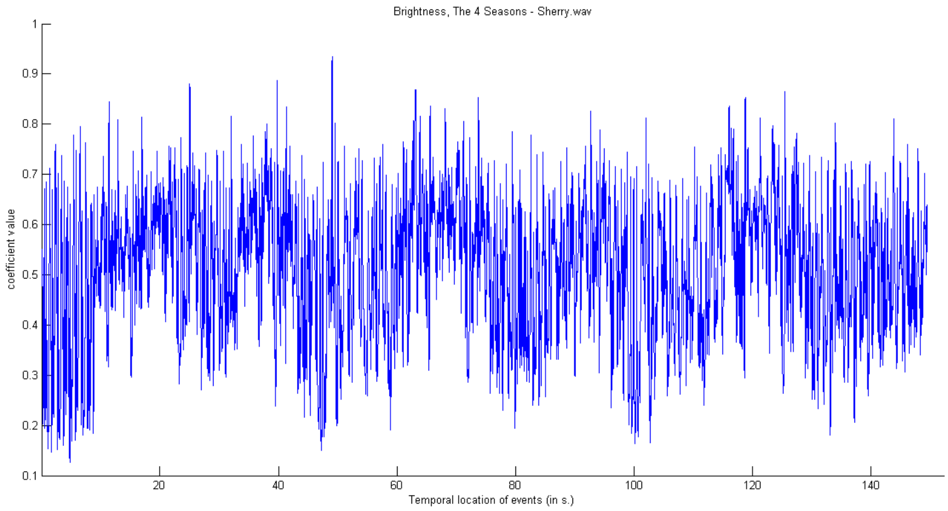


(a) ZCC computation of *The 4 Seasons - Sherry* (1962)

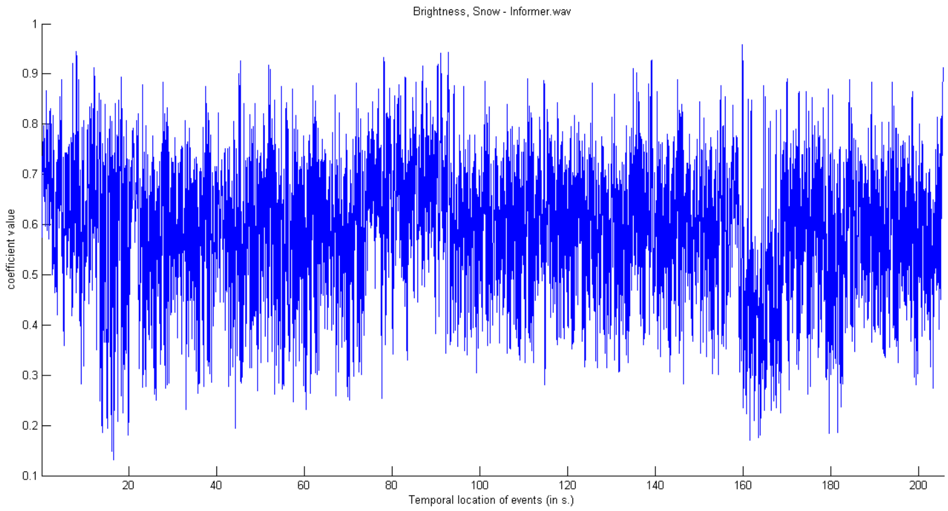


(b) ZCC computation of *Snow - Informer* (1993)

Figure D.2: Comparison of ZCC values along time for two different songs. They are computed for frames of 50ms and half overlapping.



(a) Brightness computation of *The 4 Seasons* - *Sherry* (1962)



(b) Brightness computation of *Snow* - *Informer* (1993)

Figure D.3: Comparison of brightness evolution for two different songs. It is computed for frames of 50ms and half overlapping.

Appendix E

Evaluation measures used

In pattern recognition, information retrieval and binary classification, precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances, while recall (also known as sensitivity) is the fraction of relevant instances that have been retrieved over total relevant instances in the image [54]. Accuracy is a description of systematic errors, a measure of statistical bias; as these cause a difference between a result and a *true* value. F-measure (also F-score or F_1score) is a measure of a test's accuracy.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{E.1}$$

$$Precision = \frac{tp}{tp + fp} \tag{E.2}$$

$$Recall = \frac{tp}{tp + fn} \tag{E.3}$$

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{E.4}$$

Where tp (true positives), fp (false positives), tn (true negatives) and fn (false negatives) are computed comparing the classified label with the original one (perceptually classified). For classification tasks, the terms true positives, true negatives, false positives, and false negatives compare the results of the classifier under test with trusted external judgments (perceptual classification of the data-set). The terms positive and negative refer to the classifier's prediction (sometimes known as the expectation), and the terms true and false refer to whether that prediction corresponds to the external judgment (sometimes known as the observation).

Accuracy $\in$ [0,1], Precision $\in$ [0,1], Recall $\in$ [0,1] and F-measure $\in$ [0,1]. Note that for F-measure = Precision = Recall = Accuracy = 1 a perfect classification is achieved.

Bibliography

[1] Jean-Julien Aucouturier and Emmanuel Bigand. Mel cepstrum & ann ova: The difficult dialog between mir and music cognition. In Fabien Gouyon, Perfecto Herrera, Luis Gustavo Martins, and Meinard Müller, editors, *ISMIR*, pages 397–402. FEUP Edições, 2012.

[2] Billboard. Billboard charts archive. http://www.billboard.com/archive/charts, 2017. [Online; accessed 08-May-2017].

[3] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, pages 144–152, New York, NY, USA, 1992. ACM.

[4] Eric Cai. Machine learning lesson of the day – overfitting and underfitting. http://www.statsblogs.com/2014/03/20/machine-learning-lesson-of-the-day-overfitting-and-underfitting/, 2014. [Online; accessed 20-June-2017].

[5] Kevin Chong. *The Thrill Is Gone.* Canadian Broadcasting Corporation, aug 2006.

[6] N. Cristianini and J. Shawe-Taylor. An introduction to support vector machines. http://www.support-vector.net/, 2000.

[7] Universitat Politècnica de Catalunya (UPC). Acords de distribució de programari per a l'estudiantat. https://serveistic.upc.edu/ca/distsoft/el-servei/programari-per-a-estudiantsi-professors, 2017. [Online; accessed 15-June-2017].

[8] Sophie Donnadieu. Mental representation of the timbre of complex sounds. In James W. Beauchamp, editor, *Analysis, Synthesis, and Perception of Musical Sounds*, chapter 8, pages 272–319. Springer New York, New York, NY, 2007.

[9] Taffeta M. Elliott, Liberty S. Hamilton, and Frédéric E. Theunissen. Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *The Journal of the Acoustical Society of America*, 133(1):389–404, 2013.

[10] Dan Ellis. Using knowledge to organize sound: The prediction-driven approach to computational auditory scene analysis, and its application to speech/nonspeech mixtures, 1998.

[11] Daniel P. W. Ellis. PLP and RASTA (and MFCC, and inversion) in Matlab, 2005. online web resource.

[12] A. Eronen. Comparison of features for musical instrument recognition. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, pages 19–22, 2001.

[13] Ichiro Fujinaga. Machine recognition of timbre using steady-state tone of acoustic musical instruments, 1998.

[14] Ichiro Fujinaga and Karl MacMillan. Realtime recognition of orchestral instruments, 2000.

[15] Terrence S. Furey, Nello Cristianini, Nigel Duffy, David W. Bednarski, Michèl Schummer, and David Haussler. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16(10):906, 2000.

[16] Fabien Gouyon, François Pachet, Olivier Delerue, and Henkjan Honing. On the use of zero-crossing rate for an application of classification of percussive sounds. *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, pages 1–6, 2000.

[17] John M. Grey. *An exploration of musical timbre*. Dept. of Music, Stanford University, 2 edition, 1975.

[18] John M. Grey. Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, 61(5):1270–1277, 1977.

[19] John M. Grey and John W. Gordon. Perceptual effects of spectral modifications on musical timbres. *The Journal of the Acoustical Society of America*, 63(5):1493–1500, 1978.

[20] Douglas M. Hawkins. The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, 44(1):1–12, 2004. PMID: 14741005.

[21] M. Ito and R. Donaldson. Zero-crossing measurements for analysis and recognition of speech sounds. *IEEE Transactions on Audio and Electroacoustics*, 19(3):235–242, September 1971.

[22] Jones IV and James T. Mainstream rap; cutting-edge sound tops pop in a year of controversy; video's child take beat to new streets. *USA Today*, December 1990.

[23] Paul Iverson and Carol L. Krumhansl. Isolating the dynamic attributes of musical timbre. *The Journal of the Acoustical Society of America*, 94(5):2595–2603, 1993.

[24] C. Joder, S. Essid, and G. Richard. Temporal integration for audio classification with application to musical instrument classification. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(1):174–186, Jan 2009.

[25] C. Joder, S. Essid, and G. Richard. Temporal integration for audio classification with application to musical instrument classification. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(1):174–186, Jan 2009.

[26] Patrik N Juslin. Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology*, 26(6):1797–1813, 2000.

[27] Patil K., Pressnitzer D., Shamma S., and Elhilali M. Music in our ears: The biological bases of musical timbre perception. *PLoS Computational Biology*, 8(11), nov 2012.

[28] Carol L. Krumhansl. Why is musical timbre so hard to understand? *Structure and Perception of Electroacoustic Sound and Music*, pages 43–53, 1989.

[29] Olivier Lartillot. Mirtoolbox. https://es.mathworks.com/matlabcentral/fileexchange/24583-mirtoolbox, 2017. [Online; accessed 15-April-2017].

[30] Beth Logan. Mel frequency cepstral coefficients for music modeling. *Cambridge Research Laboratory. Compaq Computer Corporation*, 2000.

[31] Brian Longhurst. *Popular Music and Society*. Polity, 2 edition, 2007.

[32] Jack Lule. The evolution of popular music. In *Understanding media and culture: An introduction to mass communication*, chapter 6.2, pages 207–223. University of Minnesota Libraries Publishing, 2016.

[33] Keith Dana Martin. Sound-source recognition: A theory and computational model, 1999.

[34] Matthias Mauch, Robert M. MacCallum, Mark Levy, and Armand M. Leroi. The evolution of popular music: Usa 1960–2010. *Royal Society Open Science*, 2(5), 2015.

[35] Stephen McAdams. Musical timbre perception. In Diana Deutsch, editor, *The Psychology of Music*, chapter 2, pages 35–67. Academic Press, 32 Jamestown Road, London NW1 7BY, UK, 3 edition, October 2012.

[36] Stephen McAdams and Emmanuel Bigand. Thinking in sound: The cognitive psychology of human audition. *Music Perception: An Interdisciplinary Journal*, 12(3):367–375, 1995.

[37] Stephen McAdams, Suzanne Winsberg, Sophie Donnadieu, Geert De Soete, and Jochen Krimphoff. Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58(3):177–192, 1995.

[38] N. Merhav and C.-H. Lee. On the asymptotic statistical behavior of empirical cepstral coefficients. *Trans. Sig. Proc.*, 41(5):1990–1993, May 1993.

[39] Nell Irvin Painter. *Creating Black Americans: African-American History and Its Meanings, 1619 to the Present.* Oxford: Oxford University Press, 2006.

[40] Geoffroy Peeters, Bruno L. Giordano, Patrick Susini, Nicolas Misdariis, and Stephen McAdams. The timbre toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, 130(5):2902–2916, 2011.

[41] DK Peneny. The history of rock and roll. http://www.history-of-rock.com/. [Online; accessed 04-April-2017].

[42] J. Pons, O. Slizovskaia, R. Gong, E. Gómez, and X. Serra. Timbre analysis of music audio signals with convolutional neural networks. *ArXiv e-prints*, mar 2017.

[43] Plomp R. Timbre as a multidimensional attribute of complex tones, 1970.

[44] Miles Raymer. Who owns trap? *Chicago Reader*, November 2012.

[45] Joan Serrà, Alvaro Corral, Marián Boguñá, Martín Haro, and Josep Lluís Arcos. Measuring the evolution of contemporary western popular music. *CoRR*, abs/1205.5651, 2012.

[46] John Shepherd. *Continuum Encyclopedia of Popular Music of the World: Volume II: Performance and Production.* Continuum, 1 edition, 2003.

[47] Kai Siedenburg, Ichiro Fujinaga, and Stephen McAdams. A comparison of approaches to timbre descriptors in music information retrieval and music psychology. *Journal of New Music Research*, 45(1):27–41, 2016.

[48] Janice C. Simpson. Music: Yo! rap gets on the map. *Time*, February 1990.

[49] B. L. Sturm. A simple method to determine if a music information retrieval system is a "horse". *IEEE Transactions on Multimedia*, 16(6):1636–1644, Oct 2014.

[50] Bob L. Sturm. Classification accuracy is not enough. *Journal of Intelligent Information Systems*, 41(3):371–406, 2013.

[51] David P. Szatmary. *Rockin' in Time: A Social History of Rock-And-Roll*. Prentice Hall, 7 edition, 2010.

[52] David L. Wessel. Timbre space as a musical control structure. *Computer Music Journal*, 3(2):42–52, June 1979.

[53] Wikipedia. Disco - wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Disco, 2017. [Online; accessed 06-April-2017].

[54] Wikipedia. Precision and recall - wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Precision_and_recall, 2017. [Online; accessed 26-June-2017].

[55] Wikipedia. Punk rock - wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Punk_rock, 2017. [Online; accessed 06-April-2017].

[56] Wikipedia. Reggaeton - wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Reggaeton, 2017. [Online; accessed 06-April-2017].

[57] Wikipedia. Trap music - wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Trap_music, 2017. [Online; accessed 07-April-2017].