

UNIVERSITY NAME

DOCTORAL THESIS

---

**A content-aware interactive explorer of  
digital music collections: The Phonos  
music explorer**

---

*Author:*

John SMITH

*Supervisor:*

Dr. James SMITH

*A thesis submitted in fulfilment of the requirements  
for the degree of Doctor of Philosophy*

*in the*

Research Group Name

Department or School Name

February 2015

*“Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism.”*

Dave Barry

UNIVERSITY NAME (IN BLOCK CAPITALS)

# *Abstract*

Faculty Name

Department or School Name

Doctor of Philosophy

**A content-aware interactive explorer of digital music collections: The  
Phonos music explorer**

by John SMITH

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

# *Acknowledgements*

The acknowledgements and the people to thank go here, don't forget to include your project advisor...

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>Contents</b>	<b>iv</b>
<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>vii</b>
<b>Abbreviations</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The importance of music analysis . . . . .	1
1.2 Phonos Project . . . . .	3
1.3 GiantSteps . . . . .	3
1.4 Purpose of this work . . . . .	5
1.5 Introduction to the problem of Playlist Generation . . . . .	5
1.6 Structure of the dissertation . . . . .	6
<b>I Background</b>	<b>9</b>
<b>2 Music Analysis Techniques</b>	<b>11</b>
2.1 Metadata . . . . .	11
2.1.1 Vector Space Model . . . . .	13
2.1.2 Co-Occurence Analysis . . . . .	13
2.1.3 Frequent Pattern Mining . . . . .	13
2.2 Audio Content Analysis . . . . .	13
2.2.1 Low-level Descriptors . . . . .	15
2.2.1.1 MFCC . . . . .	15
2.2.2 Mid-level Descriptors . . . . .	15
2.2.2.1 Rhythm . . . . .	15
2.2.2.2 Tonality . . . . .	18
2.2.3 High-level Descriptors . . . . .	20
2.2.4 Main Tools For Extracting Audio Content . . . . .	20

---

2.3	Computing Music Similarity with Audio Content Descriptors . . . . .	23
2.3.1	Notable studies on large databases . . . . .	23
2.4	Conceptual Differences Between Metadata and Audio Content Information	23
<b>3</b>	<b>Assessing the performance of a music similarity computation system</b>	<b>25</b>
3.1	Literature Review . . . . .	25
<b>II</b>	<b>Methodology</b>	<b>26</b>
<b>4</b>	<b>Case Study: Requirements and Approach</b>	<b>28</b>
4.1	Catalogue of music . . . . .	28
4.2	Requirements . . . . .	28
4.3	Design of the system . . . . .	29
<b>5</b>	<b>Off-line computation of audio features</b>	<b>31</b>
5.1	Audio content features extraction . . . . .	31
5.2	FastMap computation . . . . .	34
<b>6</b>	<b>The Real-Time Application</b>	<b>36</b>
6.1	Implementation (python server + html client), Gstreamer . . . . .	36
6.2	Functioning . . . . .	36
<b>III</b>	<b>Results and Discussion</b>	<b>37</b>
<b>7</b>	<b>Evaluation</b>	<b>39</b>
<b>8</b>	<b>Future Work</b>	<b>40</b>
<b>A</b>	<b>List of Essentia descriptors</b>	<b>41</b>
<b>B</b>	<b>List of Echonest Features</b>	<b>44</b>
<b>C</b>	<b>Phonos: list of songs</b>	<b>45</b>
	<b>Bibliography</b>	<b>62</b>

# List of Figures

1.1	Phonos . . . . .	3
1.2	GiantSteps . . . . .	5
4.1	Test . . . . .	30
5.1	Schema for the extraction of audio features. . . . .	32

# List of Tables

2.1	Main tonal descriptors . . . . .	19
5.1	List of descriptors computed offline . . . . .	34
A.1	List of features computable with Essentia . . . . .	43
B.1	List of audio features provided by Echonest . . . . .	44
C.1	Phonos catalogue of songs . . . . .	61



# Abbreviations

<b>MFCC</b>	List Abbreviations Here
<b>GUI</b>	List Abbreviations Here
<b>HPCP</b>	List Abbreviations Here
<b>MIR</b>	List Abbreviations Here
<b>RS</b>	Recommender Systems
<b>BPM</b>	Recommender Systems
<b>FFT</b>	Recommender Systems

*For/Dedicated to/To my...*

# Chapter 1

## Introduction

### 1.1 The importance of music analysis

The incredible growth of the Web over the last pair of decades has drastically changed many of our habits. One of the areas that have been highly affected by this fast-paced growth is our consumption of multimedia contents: the use of physically-stored content is seeing itself heavily reduced, as we are more and more getting used to the access of huge databases of multimedia content through the Web.

(it would be nice to cite [3] here to present the subject in a more elegant way) Music is one of these fields that have been revolutionized by this trend: the last decade has seen the rise of several Web services (iTunes, Spotify, Pandora, Google Music just to name a few) that offer their users an easy way to access their enormous catalogue of songs. Statistics show an increasing rate of annual growth for each of these services, in both the amount of users and of revenues: now they are among the most used ways of enjoying and discovering music.

However, the transition to this type of services has brought to some new problems. One of them relies on the vastness of these databases: given that users want to easily discover new music suitable to their tastes through intelligently created playlists, a way to reasonably pick songs and artists among the entire catalogue is needed.

This, among others, has been one reason of the rapid growth of **Music Information Retrieval** (MIR), an interdisciplinary research field whose subject is to provide new ways of finding information in music. Main techniques of describing music can be grouped into two categories:

- Metadata (literally *data describing data*), descriptors of music not directly retrieved from the audio signal but instead from external sources <sup>1</sup>
- Audio content descriptors, automatically computed from audio.

When it comes to choosing one method over the other, it becomes clear that both these categories of tools have their own pros and cons. Regarding metadata, major concerns arise from the questionable consistency of the descriptors among the entire catalogue of music, given that they may have been extracted from several sources. Other concerns also arise from how well they actually describe the audio track. On the other hand, audio content descriptors (especially the low-level ones) may have no musical meaning and therefore they could be hard to understand. Many efforts have been taken in order to improve the methods of information extraction of both these categories. In general, however, audio content descriptors are thought to be more flexible, since they can be easily and equally computed for any track. One advantage of this technique relies on the fact that these kind of descriptors could easily be computed not just for each kind of song, but also for any segment inside of it. This has for example been exploited by *Shazam*, a widely-used smartphone app for music identification that analyzes peaks in the frequency-time spectrum throughout all song length to build a very robust song identification system [1]. Another popular product that performs audio content analysis just for short segments of a song is The Infinite Jukebox<sup>2</sup>, a web-application using *The Echo Nest* library and written by Paul Lamere, that allows users to indefinitely listen to the same song, with the playback automatically jumping to points that sound very similar to the current one. The Infinite Jukebox can be considered an application of the so-called *creative-MIR* [6], an emerging area of activity inner to MIR whose subject is to exploit MIR techniques for creative purposes. Other relevant software that exploits Echonest library for similar purposes is Autocanonizer<sup>3</sup> and Wub Machine<sup>4</sup>. However, there aren't many commercial or research-based software tools that exploit this kind of techniques for creative interaction or manipulation of audio tracks at the moment. Probably the most relevant commercial system is Harmonic Mixing Tool<sup>5</sup>, that performs audio content analysis on the user's music collection in order to allow a pleasant and harmonic fade when mixing between songs. More recently, the research-based software

---

<sup>1</sup>There is a lack of agreement on the use of the term metadata, therefore its meaning could be different in other resources. For instance, it may be used to indicate all the data describing an audio file, including the ones derived from some computation on the audio signal itself.

<sup>2</sup><http://infinitejuke.com>

<sup>3</sup><http://static.echonest.com/autocanonizer>

<sup>4</sup><http://thewubmachine.com>

<sup>5</sup>[http://www.idmt.fraunhofer.de/en/Service\\_Offerings/products\\_and\\_technologies/e\\_h/harmonic\\_mixing\\_tool.html](http://www.idmt.fraunhofer.de/en/Service_Offerings/products_and_technologies/e_h/harmonic_mixing_tool.html)

AutoMashUpper has been developed with the intent of automating generating multi-song mashup<sup>6</sup> while also allowing the user a control over the music generated [7]. WRITE MORE ABOUT AUTOMASH HERE

## 1.2 Phonos Project

Phonos project<sup>7</sup> is an initiative of the **Music Technology Group** (Universitat Pompeu Fabra, Barcelona) in collaboration with **Phonos Foundation**. Phonos was founded in 1974 by J.M. Mestres Quadreny, Andres Lewin-Richter and Luis Callejo, and for many years it has been the only studio of electroacoustic music in Spain. Many of the electroacoustic musicians in Spain attended the courses of the composer Gabriel Brncic at Phonos. It became Phonos Foundation 1982 and in 1984 it was registered at the Generalitat de Catalunya. In 1994, an agreement of co-operation with Music Technology group was established, with the purpose of promoting cultural activities related to research in the music technology. In 2014, an exhibition at Museum de la Musica has been planned, with the purpose of celebrating the 40th anniversary of Phonos and showing many of the instruments used in the studio, while allowing visitors to listen to the music works produced there during all these years. Given the songs' average length and their complexity, a way for the visitors to quickly and nicely explore a catalogue of songs produced in these 40 years was needed.



---

FIGURE 1.1: Phonos Logo.

## 1.3 GiantSteps

GiantSteps<sup>8</sup> is a STREP project coordinated by JCP-Consult SAS in France in collaboration with the MTG funded by the European Commission. The aim of this project

---

<sup>6</sup>A mashup is a composition made of two or more different songs playing together.

<sup>7</sup><http://phonos.upf.edu/>

<sup>8</sup><http://www.giantsteps-project.eu/>

is to create the "seven-league boots" for music production in the next decade and beyond, that is, exploiting the latest fields in the field of MIR to make computer music production easier for anyone. Indeed, despite the increasing amount of software and plugins for computer music creation, it's still considered very hard to master these instruments and producing songs<sup>9</sup> because it requires not only musical knowledge but also familiarity with the tools (both software and hardware) that the artist decide to use, and whose way of usage may greatly vary between each other. The GiantSteps project targets three different directions:

- Developing **musical expert agents**, that could provide suggestions from sample to song level, while guiding users lacking inspiration, technical or musical knowledge
- Developing improved **interfaces**, implementing novel visualisation techniques that provide meaningful feedback to enable fast comprehensibility for novices and improved workflow for professionals.
- Developing **low complexity algorithms**, so that the technologies developed can be accessible through low cost portable devices.

Started on November 2013, GiantSteps will last 36 months and the institutions involved are:

- **Music Technology Group**, Universitat Pompeu Fabra, Barcelona, Spain
- **JCP-Consult SAS**, France
- **Johannes Kepler Universität Linz**, Austria
- **Red Bull Music Academy**, Germany
- **STEIM**, Amsterdam, Netherlands
- **Reactable Systems**, Barcelona, Spain
- **Native Instruments**, Germany

---

<sup>9</sup> "Computer music today is like piloting a jet with all the lights turned off." (S. Jordà). <http://vimeo.com/28963593>



---

FIGURE 1.2: GiantSteps Logo.

## 1.4 Purpose of this work

The purpose of this work is to develop a software to be used by visitors during the exhibition *Phonos, 40 anys de música electrònica a Barcelona* and that allows users to easily explore a medium-sized collection of music. This software is intended to exploit latest MIR findings to create a flow of music, composed of short segments of each song, concatenated in a way that the listener can barely realize of the hops between different songs. The application must also allow users to interact with it in order to have some control over the generation of the playlist; specifically, the user should be able to give a general direction to this flow (through some sliders or others GUI elements) in regards to some relevant music features, in a way that the change in the musical output can be perceived. The application developed is meant to be part of the GiantSteps project and therefore should follow the three guidelines explained in the previous page. In addition to this, given its future use on a public place, the application is required to be easy to use also for non-musicians, as many of the visitors of the exhibition could be.

## 1.5 Introduction to the problem of Playlist Generation

The problem of playlist generation has already been addressed by many popular music platforms, such as *Last.fm*<sup>10</sup>, *Pandora*<sup>11</sup> and *Musicoverly*<sup>12</sup>. The main objective of such services is to help users find tracks or artists that are unknown to them and

---

<sup>10</sup><http://last.fm>

<sup>11</sup><http://www.pandora.com>

<sup>12</sup><http://musicoverly.com>

that they may like, providing *personalized radio playlists*. However, a playlist may be defined, in a broad way, just as a sequence of tracks [16] and therefore its use could be more general. For instance, a common use of the term playlist refers to the broadcasting radio playlists, i.e. playlists made by DJs in a radio stations and often involving popular tracks. We can therefore define the problem of playlist generation as follows [16]:

Given (1) a pool of tracks, (2) a background knowledge database, and (3) some target characteristics of the playlist, create a sequence of tracks fulfilling the target characteristics in the best possible way.

This task is made particularly challenging by the average size of the music database on which the generation of the playlist is needed: already, personal music libraries can be huge [17], hence the corresponding amount of information to be processed in order to generate the playlist leads to very heavy computational tasks. Depending on the need of the application, these tasks may also be performed offline, although a real-time user interaction should be supported in many cases in order to allow the user to have some control over this generation process (such as in the case study of this work). As we will see in Chapter 2, extracting information from an audio signal is not a trivial task and many algorithms have considerable time-complexity, and this may lead to very long computational times already for the analysis of small-sized catalogues. Playlist generation is a well-known problem inside MIR [18] [19], since this task can be considered as a retrieval task if its definition is limited to the selection of tracks satisfying a user query [16]. Other major topics of MIR also include extraction of features and similarity analysis, that can be seen as a basis for building a playlist generation system [20].

## 1.6 Structure of the dissertation

This dissertation is organized as follows:

- The first part will at first give an overview regarding music analysis techniques, explaining *metadata*, audio content analysis and the differences between them. Then, common techniques of music similarity computation will be explained.
- The second part will be about the methodology, explaining the different stages of the development, the problems faced and the techniques used. A presentation of the case study will introduce to an explanation of the reasons that lead to prefer the use of some techniques over others.



- 
- Finally, experimental results will be shown, together with some ideas regarding future development of the application.



## Part I

# Background



## Chapter 2

# Music Analysis Techniques

The main subject of MIR regards the *extraction and inference of musically meaningful features, indexing of music* (through these features) and the development of *search and retrieval schemes* [2]. In other terms, the main target of MIR is to make all the music over the world easily accessible to the user [2]. During the last two decades, several approaches have been developed, which mainly differ in the music perception category of the features they deal with. These categories generally are: *music content*, *music context*, *user properties* and *user context* [4]. *Music content* deals with aspects that are directly inferred by the audio signal (such as melody, rhythmic structure, timbre) while *music context* refers to aspects that are not directly extracted from the signal but are strictly related to it (for example label[23], artist and genre information [24] [25], year of release [26], lyrics [27] and semantic labels). Regarding the user, the difference between *user context* and *user properties* lies on the stability of aspects of the user himself. The former deals with aspects that are subject to frequent changes (such as mood or social context), while the latter refers to aspects that may be considered constant or slowly changing, for instance his music taste or education [4].

In this chapter, we will focus on the differences between the categories *music content* and *music context*.

## 2.1 Metadata

By metadata we mean all the descriptors about a track that are not based on the *music content*. Therefore, they are not directly extracted from the audio signal but rather from external sources. They began to be deeply studied since the early 2000s, when first doubts about an upper threshold of the performance of audio content analysis systems arised [5]. Researchers then started exploring the possibility of performing

retrieving tasks on written data that is related to the artist or to the piece.

At first, the techniques were adapted from the Text-IR ones, but it was immediately clear that retrieving music is fairly more complex than retrieving text, because the music retrieved should also satisfy the musical taste of the user who performed the query.

The techniques used in this category may differ both in the sources used for retrieving data and in the way of computing a similarity score, and clearly the performance of a system using metadata for similarity computation is highly affected by both of these factors. Sources may include [12]:

- Manual annotation: description provided by experts; they may be referred to genre, mood, instrumentation, artist relations.
- Collaborative filtering data: data indirectly provided by users of web communities, in the form of user ratings or listening behaviour information.
- Social tags: data directly provided by users of social network of music (such as *Last.fm*<sup>1</sup>) or social games.
- Information automatically mined from the Web. Sources in these cases may include web-pages related to music or microblogs (for instance the very popular Twitter).

The availability of some of them greatly depends on the size of the music collection under consideration; for instance, as manual expert annotations might be very accurate, they would be extremely costly and probably infeasible on large collections [8]. In contrast, collaborative filtering data may be the most studied technique, given that it may be applied to other different fields (such as movies or books recommendation) with just little changes. It is the predominant approach in the field of Recommender Systems (RS) [35] and is mainly focused on user ratings, generally leading to better results [11]. However, some concerns are related to this technique. First, collaborative filtering approaches have not been designed to be used for playlist generation, but mainly for recommending artists or music. Second, the availability of datasets for user ratings in the field of music is very limited compared to other fields, and research is often based on very small samples [36]. Regarding listening behaviour information, they might be inaccurate since they don't keep track of song durations and of the user activities while listening to music [38]. In addition, there's no way of collecting negative feedback (*dislikes*) through them and, more in general, listening to a specific song doesn't necessarily imply liking that song [12].

Sources are picked also in relation to the subject of the research or of the system, that may be for example a recommendation or a similarity computation system. At this point,

---

<sup>1</sup><http://last.fm>

it's important to highlight the difference between the two of them: a recommendation system not only has to find similar music, but has also to take into account the personal taste of the user, and therefore it's generally considered as a basic tool to produce recommendation [22]. In any case, the terms "similarity" and "recommendation" cannot be substituted, given that a good similarity computation system doesn't necessarily equate to a good recommendation system [37]. The computation of similarity may happen through a Vector Space Model (a technique adapted from the Text-IR), co-occurrence analysis or frequent-pattern mining. In the next subsections we will briefly explain the characteristics and the performance of these techniques.

### 2.1.1 Vector Space Model

The main idea of this technique lies on building a bag-of-words representation <sup>2</sup> of a retrieved document, and then computing a term weight vector for each document. It's a frequently used technique in Text-IR (and in Computer Vision) which can safely be used when retrieving web pages related to music, in an attempt of computing similarity. One of the first work in this field [39] provided an analysis of this kind on music-related web pages retrieved with the queries (to the *Google* search engine) "artist" **music review** and "artist" **genre style**, where words such as music and review were added to improve the chances of automatically retrieving webpages related to music.

### 2.1.2 Co-Occurrence Analysis

### 2.1.3 Frequent Pattern Mining

## 2.2 Audio Content Analysis

The main idea behind this kind of analysis is to directly extract useful information, through some algorithms (or library of algorithms), from the audio signal itself. The type of content information extracted may greatly vary in relation to the need of the research, but we can mainly distinguish four categories [12]:

- *Timbral* information: related to the overall quality and color of the sound.

---

<sup>2</sup>A bag-of-words can be basically seen as an extension of a programming language "dictionary": it collects words (that sometimes may just be an abstraction of much more complex features, such as computer vision descriptors) from a document, and then computes the frequency with which each of them appears in the document. Two different documents are considered similar if they contain the same or similar words with a comparable frequency.

- *Temporal* information: related to rhythmic aspects of the composition, such as tempo or length of measures.
- *Tonal* information: directly linked to the frequency analysis of the signal and to the pitch. It can describe what notes are being played or the tonality of a given track.
- *Inferred semantic* information: information inferred (usually through machine learning techniques) from the previous categories, in the attempt of giving a more defined and understandable shape to the data collected. This kind of information may include descriptors such as genre, valence or arousal.

Information extracted through this family of techniques may also be categorized in the following way:

- Low-level data: information that has no musical meaning and that, more in general, is not interpretable by humans. Examples of this kind of descriptors are Mel Frequency Cepstral Coefficients (MFCCs) and Zero Crossing Rate (ZCR).
- Mid-level data: information that has musical meaning but that is related to low-level music features. This kind of category mainly includes temporal and tonal descriptors.
- High-level data: corresponding to inferred semantic information.

Many of the studies conducted on the computation of music similarity through audio content descriptors have solely focused on low-level and timbral information, because this has been proved to bring alone to acceptable results with proper similarity measures [13]. However, more recent studies have shown some evidence of advantages in using high-level descriptors [14] [15] and, more in general, the most performant systems use data from all of these categories. When computing low and mid-level descriptors, the procedure requires the following operations:

- Conversion of the signal from stereo to mono, in order to compute all the descriptors for just one signal
- Down-sampling of the signal to improve the performance while computing the descriptors
- Segmentation of the signal into frames, short segments (usually from 512 to 2048 audio samples). Consecutive frames are usually not disjoint: the so-called *hop-size* determines the hop of samples between the beginning of a frame and the next one, and is normally half or a quarter as big as the *frame size*.



- Computation of Fast Fourier Transform, with an appropriate prior windowing technique <sup>3</sup>.

The computation of descriptors is then performed on each frame, and finally a single value for each descriptor is computed by the means of some statistical analysis. Mean, median, variance and covariance are the most used statistical tools for calculating representative global values out of the enormous *pool* of values computed in each frame. Some more operations may sometimes be needed, such as de-noising <sup>4</sup> of time-scaling of the signal.

In the next sections, a more detailed look among most important descriptors will be given.

## 2.2.1 Low-level Descriptors

### 2.2.1.1 MFCC

## 2.2.2 Mid-level Descriptors

### 2.2.2.1 Rhythm

In traditional music notation, there are several notations for tempo. It may be expressed in BPM (beats per minute), MPM (measures per minute; commonly used in ballroom dance music) or by semantic notations indicating a range of BPM; an example of this last category of notations may be the popular system of Italian markings, such as *presto* (168-200 BPM), *andante* (84-90 BPM) or *allegro* (120-128 BPM).

In the field of MIR, accurate notations are needed, therefore semantic annotations are disregarded in favour of more precise notation such as BPM and Onset Rate (OR).

### Onset Rate

IDEAS: image of ADSR, some flowchart for a standard onset detection algorithm (look at figures folder) Onsets are generally defined as the beginning of a new musical note, and onset rate is therefore defined as the number of onsets in a time interval. This definition however hides several difficulties: in polyphonic music, nominally simultaneous notes may be spread over tens of seconds, making this definition more blurred [40]. Moreover, several instruments have a long attack time and this makes the task of defining an onset

<sup>3</sup>Although this last step may not be strictly seen as a necessary operation, many descriptors rely on frequency analysis of the signal and therefore they require the computation of the Fourier Transform.

<sup>4</sup>A set of operations which purpose is to reduce the amount of background noise in a signal, therefore incrementing the signal-to-noise ratio (*SNR* or sometimes *S/N*).

time even harder.

Several ways of computing an onset detection function have been proposed, according to what aspects are taken into account for defining an onset. Actually, onset detection may be performed in time domain (when looking for significant changes in the overall energy), frequency domain (if looking for events regarding just a specific range of frequencies), phase domain or complex domain. Important algorithms for this task are:

- *HFC*, the High Frequency Content detection function that looks for important changes on highest frequencies. It is very useful for detecting percussive events.
- Spectral Flux, that decomposes the entire audible range of frequencies (approximately the interval 20-20000 Hz) into bins, measures changes in magnitude in each bin, and then sums all the positive changes across all the bins.
- the Complex-Domain spectral difference function [41] taking into account changes in magnitude and phase. It emphasizes note onsets either as a result of significant change in energy in the magnitude spectrum, and/or a deviation from the expected phase values in the phase spectrum, caused by a change in pitch.

HFC was proposed by Masri in [42]. Let us consider the short-time Fourier transform (STFT) of the signal  $x(n)$ :

$$X_k(n) = \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} x(nh + m)w(m)e^{-\frac{2j\pi mk}{N}} \quad (2.1)$$

where  $w(m)$  is again an  $N$ -point window, and  $h$  is the hop size, or time shift, between adjacent windows. The idea behind HFC is to give more weight to higher frequencies, by defining an onset function whose values are computed in the following way:

$$HFC(n) = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} k |X_k(n)|^2 \quad (2.2)$$

The HFC function produces sharp peaks during attack transients and is notably successful when faced with percussive onsets, where transients are well modeled as bursts of white noise [43].

On the other hand, the Spectral Flux  $SF$  function is defined as follows:

$$SF(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} H(|X(n, k)| - H(|X(n-1, k)|)) \quad (2.3)$$

where  $H = \frac{x+|x|}{2}$  is the half-wave rectifier function. This algorithm greatly characterizes changes in magnitude spectrum but it is quite weak to frequency-modulation phenomena (such as vibrato). To this end, the recently proposed variant SuperFlux [44] seems to achieve much better results.

Another interesting onset function is the *Complex Domain*, that calculates expected the expected amplitude and phase of the current bin  $X(n, k)$  based on the previous two bins  $X(n-1, k)$  and  $X(n-2, k)$ . By assuming constant amplitude the expected value  $X_T(n, k)$  is then computed:

$$X_T(n, k) = |X(n-1, k)|e^{j\psi(n-1, k)+\psi'(n-1, k)} \quad (2.4)$$

and therefore a complex domain onset detection function CD can be defined as the sum of absolute deviations from the target values [40]:

$$CD(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |X(n, k) - X_T(n, k)| \quad (2.5)$$

Given an onset function (for instance one of the already cited  $HFC(n)$ ,  $SF(n)$  or  $CD(n)$ ), onsets are then extracted by a peak-picking algorithm which finds local maxima in the detection function, subject to various constraints. Threshold and constraints used in the peak-picking algorithm has a large impact on the results, specifically on the ratio of false positives<sup>5</sup> to false negatives<sup>6</sup>. For instance, a higher threshold may lead to a lower number of false negatives but to a higher number of false positive, while a lower threshold may have the opposite effect. A compromise, mostly specific to the application, has to be found.

## BPM

The algorithms for detecting the beats-per-minute (generally called *beat tracking algorithms*) greatly rely on onset detection functions. The basic idea is to look for some time-pattern that may explain the distribution of onsets over time, and hence derive BPM. They usually require more than one onset detection function to achieve good results. One of the most performant beat tracking algorithm is TempoTapDegara, presented by N. Degara et al. in [45].

## EXPLAIN ALGORITHM HERE

<sup>5</sup>True onsets that are not detected by the algorithm.

<sup>6</sup>Points that are classified as onsets by the algorithm, while they are actually not.

### 2.2.2.2 Tonality

Many efforts have been taken in order to improve the techniques for detecting tonality or harmonic content of a song, as this is one of the most main aspects of western music (a direct consequence of tonality is the detection of predominant melody; to understand why this is so important just ask yourself how many times you whistled or sang a song to let other people recognize it). Many studies have focused on this aspect of music were not oriented toward the computation of similarity between tracks, but instead toward different tasks, such as *automatic trascription of a polyphonic audio signal* (mainly into a MIDI representation) and *source separation*, that is the task of isolating a single and specific instrument among many playing together.

From a music point of view, in western music, an octave is made of 12 different pitches, and seven different notes take place in this discrete range. According to the pitch assigned to each note, we may have different *keys*, that are a combination of a *tonic* (the central pitch) and the mode. *Major* and *minor* are the most popular modes. (ADD IMAGE OF THESE TWO MODES)

*Harmony* is a term that denotes the simultaneous combination of notes, called *chords*, and over time, *chord progressions*.

One of the most important descriptor for extracting information related to tonality is called Harmonic Pitch Content Profile (***HPCP***, also called chromagram). This is directly related to tonality and chord detection: chords can be recognized from from the *HPCP* without even precisely detecting what notes are being played, and tonality can also be inferred by *HPCP* (and in this case a previous estimation of chords is not necessary).

An *HPCP* is a  $12k$  size vector indicating the level energy for each profile class. If  $k = 1$ , the *HPCP* represents the intensities of the twelve semitone pitch classes, otherwise of subdivision of these<sup>7</sup>. In [47], Gómez proposes to distinguish tonality features on temporal scales:

- Instantaneous: features attached to an analysis frame.
- Global: features related to a wider audio segment, for instance a phrase, a chorus or the whole song.

Furthermore, Gómez proposes to split tonal descriptors in both low-level and high-level descriptors. We hence obtain the representation of tonal descriptors shown in Table 2.1.

---

<sup>7</sup>It may be extremely useful to study subdivision of semitone pitch classes when dealing with non-chromatic scales, that are very popular in eastern music.

Name	Temporal Scale	Level of abstraction
HPCP	Instantaneous	Low
Chord	Instantaneous	High
Average HPCP	Global	Low
Key	Global	High

TABLE 2.1: Main tonal descriptors.

The general approach for computing *HPCP* is indicated in figure PUT BLOCK DIAGRAM HERE and can be summarized as follows:

- At first, some pre-processing of the audio signal may be performed. For instance, a transient detection algorithm may be used to detect and eliminate regions where the harmonic structure is noisy. This step is usually performed to decrease the computational cost of the *HPCP* without affecting its output [48].
- At this point, spectral analysis is required: once the signal is segmented into frames of a proper size and a windowing function is applied, the Fast Fourier Transform (*FFT*) is computed to get the frequency spectrum. Frames should not be too short, in order to have a better frequency resolution.
- A peak-picking algorithm is run to find those frequencies corresponding to local maxima in the spectrum. Usually, these algorithms are not run only on the interval [100, 5000] Hz: this has shown much better results, because outside this range the signal is predominantly noisy, due to some percussion and instrumental noise [47].
- The *HPCP* is finally computed; many approaches have been developed for this task, all based on the pitch content profile algorithm presented by Fujishima in [46]. At first, a mapping of each frequency bin of the *FFT* to a pitch class is needed (for instance, *FFT* bins corresponding to frequencies like 430Hz, 432Hz or 444Hz are mapped to the A at 440Hz). Then, the amplitudes inside each region are summed up and divided by the number of bins inside that region. Finally, the bins obtained are collapsed, so that bins referring to the same note but in a different octave (for example A4 and A5) are collapsed in a single bin for that note, indicating the overall energy of it in the frame. The *HPCP* is different from the PCP in the sense that a higher resolution may be used for *HPCP* bins (decreasing the quantization level to less than a semitone) and a weight function is introduced into the feature computation. The *HPCP* value of the  $n$ -th *HPCP* bin is calculated as:

$$HPCP(n) = \sum_{i=1}^{nPeaks} w(n, f_i) a_i^2 \quad (2.6)$$

where  $a_i$  and  $f_i$  are respectively the magnitude and the frequency of the  $i$ th peak,  $nPeaks$  is the number of spectral peaks considered, and  $w(n, f_i)$  is the weight of the frequency bin  $f_i$  when considering the *HPCP* bin  $n$ .

The performance of the *HPCP* builder strongly relies on the weight function [49]. Note that, for how the common procedure of building *HPCP* is defined, *HPCP* are usually considered robust to noise and different tuning references.

*HPCP* values are usually normalized in order to store the relative importance of the  $n$ th *HPCP* bin:

$$HPCP_{normalized}(n) = \frac{HPCP(n)}{\max_n(HPCP(n))} \quad (2.7)$$

Once the *HPCP* are computed, additional tonal features may be computed, such as tonality or chords. Regarding tonality estimation, this is generally computed through a correlation analysis between the *HPCP* obtained and a matrix of *HPCP* profiles corresponding to different keys.

### 2.2.3 High-level Descriptors

### 2.2.4 Main Tools For Extracting Audio Content

Many tools are available for the extraction of audio content descriptors from an audio signal. Many of them have been developed by researchers following the research necessities of MIR. This great variety of tools offers support to several operative systems (mainly Linux, Mac OS X and Windows) and programming languages (Java, C++, C, Python, Matlab). Some of this tools may be offered as standalone software or as a Vamp plugin. Not all the tools for extracting audio content are open-source, therefore not being particularly useful for the research community. In the following paragraphs, we'll briefly show the features of the tools taken into account on the development of this work.

#### Essentia

Essentia<sup>8</sup> is an open-source C++ library of algorithms for audio analysis and audio-based music information retrieval. It has been developed at Music Technology Group<sup>9</sup>, Universitat Pompeu Fabra, and has released under the Affero GPL license<sup>10</sup>. In its current version 2.0.1, it contains a large collection of spectral, temporal, tonal, and

<sup>8</sup><http://essentia.upf.edu/>

<sup>9</sup><http://mtg.upf.edu/>

<sup>10</sup><http://www.gnu.org/licenses/agpl.html>

high-level music descriptors, algorithms for audio input/output functionality, standard digital signal processing blocks and statistical tools. The library can be complemented with Gaia <sup>11</sup>, a C++ library to apply similarity measures and classifications on the results of audio analysis. Both these libraries include Python 2.\* bindings and support Linux, Mac OS X and Windows. Essentia has been in developed for over 7 years, incorporating the work of more than 20 researchers and developers through its history. It offers two different modes: standard and streaming, the first being imperative while the latter being declarative. Each processing block is offered as an algorithm, and has three different types of attributes: inputs, outputs and parameters. Different blocks may be linked in order to perform the required processing task. In figure INSERT FIGURE a block diagram of a processing task is shown, composed of several different algorithms linked together. See Appendix A for consulting the full list of descriptors provided by Essentia 2.0.1.

### The Echo Nest

*The Echo Nest*<sup>12</sup> is a company that provides access, through Web API, to a collection of audio descriptors for a catalogue of over 36 million songs and almost 3 million artists. In order to access to this database, an API key is required, and some rate limits are imposed to the use of a free license (for instance, the maximum number of HTTP calls for minute is subject to a limit, generally 20). Users can decide to upload their collection into this database, so that descriptors will be computed for new songs and made available to other users. The performance of this library greatly depends on the chance that a song that is about to be analyzed has already been uploaded into this service. If this is not the case, the upload time has to be taken into account for performing the analysis task.

*The Echo Nest* provides a great amount of descriptors for each track (see appendix B for the entire list), ranging from very accurate audio content information to metadata, and has been used by several commercial solutions, developed by *Spotify*, *Rdio*, *Warner Music Group* and many others. Many official and unofficial libraries provide access to *The Echo Nest* service; among these, the most important one is probably the official Python library *Pyechonest*<sup>13</sup>, that provides full access to all of the Echo Nest methods including artist search, news, reviews, blogs, similar artists as well as methods for retrieving detailed analysis information about an uploaded track. Furthermore, the library *Echo Nest Remix*<sup>14</sup> worths mentioning, as it is a library for audio manipulation and mixing

---

<sup>11</sup><https://github.com/MTG/gaia>

<sup>12</sup><http://the.echonest.com/>

<sup>13</sup><http://echonest.github.io/pyechonest/>

<sup>14</sup><http://echonest.github.io/remix/>

and has been used by many *web-applications*, including The Infinite Jukebox.

However, the source code of *The Echo Nest* API service is not provided, therefore it has little usefulness to the research community.

*The Echo Nest* has been acquired by *Spotify* on March 2014.

## jMIR

jMir<sup>15</sup> is an open-source software suite implemented in Java and intended for use in music information retrieval research. Its development has been guided by Cory McKay (Marianopolis College), with many researchers from several institutions contributing to it. *jMir* is composed of several components differentiated in their scope, spacing from audio content analysis (performed by *jAudio*), to web mining of metadata and machine learning algorithms for classification.

The most relevant components of this suite are as follows:

- *ACE*: Pattern recognition software that utilizes meta-learning.
- *jAudio*: Software for extracting low and high-level features from audio recordings.
- *jSymbolic*: Software for extracting high-level features from MIDI recordings.
- *jWebMiner*: Software for extracting cultural features from web text
- *jSongMiner*: Software for identifying unknown audio and extracting metadata about songs, artists and albums from various web services.

## MIRtoolbox

MIRtoolbox<sup>16</sup> is a set of functions for Matlab, dedicated to the extraction of audio content features from audio files. The design is based on a modular framework: algorithms are decomposed into stages, formalized using a minimal set of elementary mechanisms, with the objective of offering an overview of computational approaches in the MIR research field. MIRtoolbox has been developed at the Jyväskylän Yliopisto (University of Jyväskylä, Finland), by Olivier Lartillot, Petri Toivainen and Tuomas Eerola.

---

<sup>15</sup><http://jmir.sourceforge.net/>

<sup>16</sup><https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox>



## 2.3 Computing Music Similarity with Audio Content Descriptors

### 2.3.1 Notable studies on large databases

When large music collections are used, performance of similarity computation algorithms become critic. Although the collection to be used by the system during its public use can't be considered large, the necessary decomposition of it into hundred of thousands excerpt to be analyzed just in few seconds makes performance a critical factor when designing and implementing an algorithm. Therefore, a deep look into studies where large collections were used was needed.

The first content-based music recommendation system working on large collections (over 200,000 songs) was published by Cano et al. in [52], 2005. The system presented on this work relies on rhythmic and timbre features, combined to form a music similarity feature vector. No special indexing technique was used.

The first music recommendation system for large databases using Gaussian timbre features was proposed some months later by Roy et al. in [53]. In this work, they propose to use a Monte-Carlo approximation of the Kullback-Leibler (KL) divergence to measure music similarity. The method proposed is, in principle, similar to the one proposed by Schnitzer et al. in [51], which has been used on the development of this thesis work (see 5.2). To pre-filter their results, Roy et al. increase the sampling rate of the Monte-Carlo approximation. As the divergence values converge, they are able to reduce the number of possible nearest neighbors. This method has shown good performance, both in query time and results.

A different attempt of improving performance was proposed by Levy and Sandler in [54] where they use only diagonal covariance matrix instead of a full one to compute music similarity. While this has shown a ten-fold speedup compared to the full Kullback Leibler divergence, the quality of this simpler similarity measure results in worse genre classification rates.

## 2.4 Conceptual Differences Between Metadata and Audio Content Information

The performance of content-based approaches is considerably lower [9]. It is challenging to try to make the so-called *semantic gap* smaller [10]

The advantage of relying on the audio signal over, say, expert annotations, is that the process is objective and can be automated to a large extent. However, extracting

the features can be computationally costly [21]. Another limitation is that there might be features like the release date, the “freshness,” or popularity of a track, which can be relevant in the playlist generation process but that cannot be extracted from the audio signal [22].

When used in an automated process, data completeness and consistency are crucial. Another potential problem is that not all types of metadata are objective, and annotations regarding, for example, the mood or the genre of a track can be imprecise or inconsistent [28].

(speaking of tags) Although such annotations can be rich and diverse, the perception of music is again subjective and can even be influenced by the perception of other people [29]; tags only for popular songs [28]

When dealing with track ratings: grabbed from a wall posting on Facebook [30] or a tweet on twitter [31], 1-to-5 rating scales like on iTunes. Challenges: problem of data sparsity (especially for the tracks from the long tail), a positivity bias (the phenomenon that most of the ratings are highly positive and negative feedback is rare [28]).

## Chapter 3

# Assessing the performance of a music similarity computation system

### 3.1 Literature Review

The coherence of the tracks is a typical quality criterion for playlists [32]. Therefore, selecting and ordering tracks based on their similarities is an obvious strategy to generate playlists. The core of any similarity-based approach is its distance function, which characterizes the closeness of two tracks. How the distance function is actually designed depends on the available data, which could include the raw audio signal along with the features that can be derived from it, but also metadata, such as the artists, the genres, playcounts, or ratings [Slaney and White 2007]. In many cases, a signature or model of each track is determined first, in which the distance function is then applied. Typical examples for such functions applied on more abstract models of a track's features are the earth-mover's distance [32], the Kullback-Leibler (KL) divergence [33], or the Euclidean distance [34].

See section 5.2 of [16] for a background on how to assess the quality of a playlist: user studies, log analysis, objective measures, comparison with handcrafted playlists

## Part II

# Methodology



## Chapter 4

# Case Study: Requirements and Approach

### 4.1 Catalogue of music

The catalogue of music provided features 584 songs, for a total length of 91 hours, 43 minutes and 35 seconds. The average length of each song is 9 minutes and 25 seconds circa. This catalogue has been provided with metadata indicating only artist, year of release and title of each song. Furthermore, all of these work can be labelled as belonging to the electro-acoustic genre, which usually indicates very abstract and arrhythmic, for which is difficult to provide semantic descriptors or tags. Given this latter feature of the music and the length of the entire catalogue, the possibility of manually annotating it with proper metadata has been soon disregarded. This collection of music has therefore represented a great chance for developing a system based on the latest findings on audio content analysis.

### 4.2 Requirements

Despite its intended use as part of the exhibition “*Phonos, 40 anys de música electrònica a Barcelona*”, the software developed should feature good flexibility to different catalogues of music, in order to be exploited as a part of the research for the GiantSteps project. This has represented a strong requirement during the development, and has induced the adoption of several descriptors that may not be particularly meaningful for the Phonos catalogue of songs, but that extend the range of possible music catalogues in which the system performance could be satisfactory. Furthermore, as a

part of a research project, the system developed should be easily extendable in other research activities, hence a modular, coherent and well-document code is preferred.

The software is intended to be used at the exhibition through an interactive kiosk: it will be available to users as a link inside a more general *webpage* containing several information regarding Phonos history and instrumentation. In addition, it must fully support touch devices, provided that this will be the only way users will be able to interact with the application.

All of these requirements have lead to the choice of developing a *web-application*.

Anyways, the interactive kiosk to be used at the exhibition was not available during the development; furthermore, its technical specification was unknown. For these reasons, it was therefore decided to develop a *two-layers system* made of the interactive kiosk machine connected (by an Ethernet cable) to a server machine. The latter one is in charge of providing and executing all the complex functions required during the functioning of the system.

Furthermore, the system must react to the real-time interaction of users with the user interface. Computation times must hence be as low as possible, in order to avoid a notable and inconvenient delay between the user interaction and the effective perception of changes in the flow of audio.

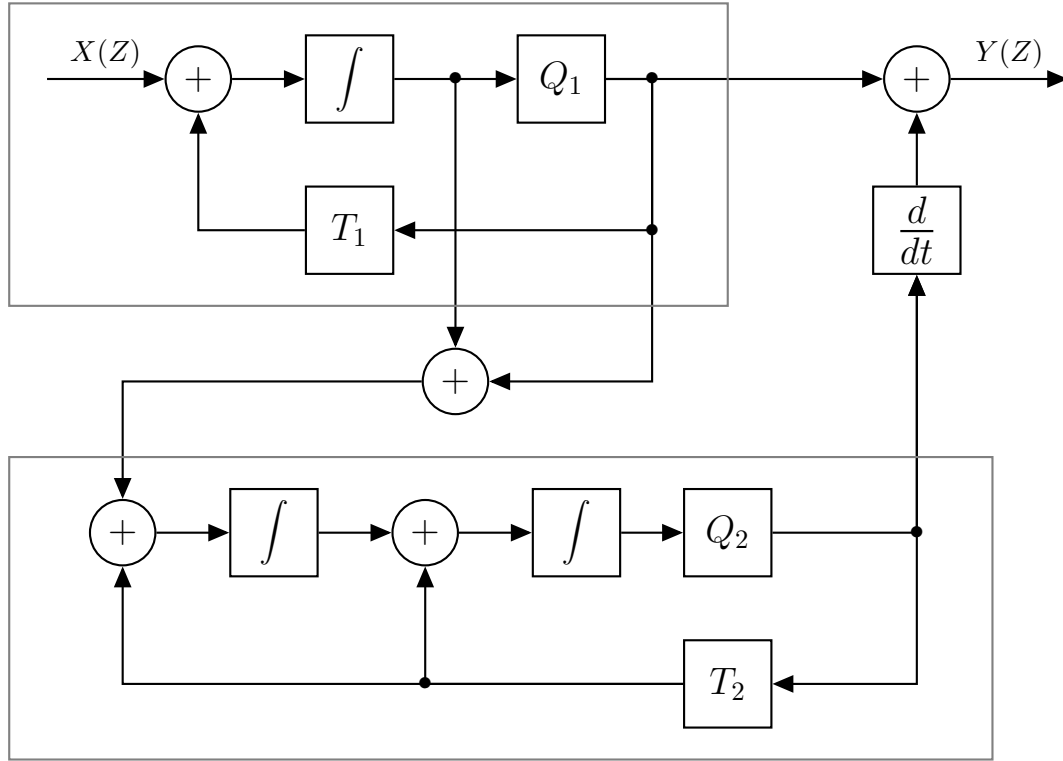
Finally, given the substantial average length of the songs, the system should segment the songs into very short excerpts (from 2 seconds to around 30; the choice of this length should be available to users in a real-time fashion), in order to allow users to listen to as many works as possible during the visit at the exhibition and to more easily find tracks that fulfill their taste. It must then be found a way to properly segment the audio pieces and computing descriptors for each slice obtained. In order to achieve a better sense of “*flow of music*”, the computation of similarities should therefore be carried out between these short excerpts, instead of exploiting descriptors for the entire songs.

### 4.3 Design of the system

The requirements cited above have lead to the following choices for the design of the system. First, the computation of audio descriptors can be performed offline, because the catalogue of music on which the system will run is not subject to changes. It is therefore safe to compute descriptors prior to the public use. The performance of the system will greatly benefit from this choice, given that the computation of audio descriptors for each excerpt of every song of the catalogue is the most computationally intensive step to be performed. The audio descriptors will be stored on the server machine.

Second, for the system is intended to have low response times to user inputs, the computation of music similarity is being carried out on the server (for the performance of the

## FIRST-ORDER NOISE SHAPER



## SECOND-ORDER NOISE SHAPER

FIGURE 4.1: Block diagram test.

interactive kiosk machine are unknown, as already cited in 4.2), with proper music similarity algorithms. The flow of music is not supposed to require an human interaction, to the meaning that it will automatically generate a flow of music based on the computation of audio similarity also without an interaction of the user. Actually, the system always concatenates segments in a way that only very similar segments are consecutive elements of the playlist. The interaction of the user will eventually give a direction to this flow, according to the user's will and taste.

Third, the application running on the server machine will be in charge of collecting the user interaction with the web-application running on the interactive kiosk machine, and that will come in the form of *HTTP POST requests*. At each user interaction, the application running on the server machine deletes the current and already computed playlist and performs an audio similarity computation between the currently playing excerpt and a set of excerpts that fulfill all the requirements about music that the user has imposed through the graphic user interface. One of the most similar excerpt is taken from the list and a new content-aware playlist starts being built above that.



## Chapter 5

# Off-line computation of audio features

In order to achieve good performance, two very computationally intensive tasks of the system are performed off-line, and their output is then going to be used by the real-time application. These tasks consist of the computation of the audio content descriptors and of the building of a *fast-map*, a high dimensionality space in which each point correspond to an excerpt. This space is built in a fashion that guarantees that nearby points of this space correspond to very similar excerpts.

### 5.1 Audio content features extraction

Solving this problem has involved two very important choices: what audio content descriptors to use and what library or tool to use for computing them.

Many factors have been taken into account for solving both of these problems.

Among the features of the tools, flexibility has constituted the strictest requirement: an easy way to compute descriptors for each excerpt of every track is required, while many tools provide only ways of computing descriptors for the entire file. In this latter case, the file should manually split into *subfiles* (one for each segment), therefore implying a huge waste of memory. This has soon lead to the exclusion of *jMir*, for it doesn't fulfill this requirement.

Second, the tool should easily be callable by source code or bash scripts, and results of the analysis must be stored in output files.

Third, the computation of descriptors should be as fast as possible, given that the excerpts to be analyzed are in the order of tens of thousands.

Last but not least, the tool must provide descriptors whose usefulness for this specific

case study has been empirically verified during the development of the system.

All of these requirements lead to the choice of performing the audio analysis with Essentia and Echo Nest: the first for its speed, flexibility and reliability. Echo Nest has been used for some of its descriptors are not present or not as accurate in Essentia, and have shown a great usefulness during the development.

Furthermore, both of the two libraries are offered in Python, allowing the entire analysis task to be written in a single programming language, therefore improving the code consistency and readability.

The schema for the extraction of the audio features is illustrated in figure 5.1.

At first, the user is required to give the path of the folder in which the audio files are

FIGURE 5.1: Schema for the extraction of audio features.

stored. The collection is entirely stored as .mp3 files with a sample rate of 44100Hz and a bitrate of 192kbps. The application then collects the path to all the .mp3 files in this folder, and mark them as to be analyzed if no previous analysis has been performed. An analysis of these files with Echo Nest (through Pyechonest) is performed, and we specifically use the following fields of the output of this analysis: *bars*, *BPM*, *loudness*, *HPCP* and *acousticness*. *Bars* give the starting and ending point of each bar detected and, although not particularly meaningful for the arhythmic Phonos catalogue of music, have shown to perform well on the additional and more generic personal catalogue used during development testing; therefore, it was decided to use them in order to improve the flexibility of the system.

Segmentation of songs into excerpts is then performed, based on starting and ending point of each bar. Then, we compute more specific descriptors with Essentia for these excerpts, with the following strategy:

- each excerpt is divided into frames, with a size of 2048 samples and a hop of 1024 samples. For each of these frames:
  - we apply an Hann windowing function
  - we apply the FFT algorithm provided by Essentia in order to get a spectral representation of the signal
  - we look for peaks in the spectrum, collecting their frequencies and magnitudes, and then we use them to compute the dissonance in the frame, with Essentia's algorithm *Dissonance*
  - an HFC onset function is computed on the spectrum, that will be used afterwards to compute the onset times

- the MFCC bands and coefficients are computed with Essentia’s algorithm `Mfcc`<sup>1</sup>
- the energy in 27 Bark bands of the spectrum is computed
- onset times in the excerpt are calculated, according to the onset function computed in each frame, and then onset rate is calculated with the formula:

$$OR_{excerpt} = \frac{Onsets_{excerpt}}{Length_{excerpt}} \quad (5.1)$$

- dissonance in the excerpt is computed as a mean of the dissonance in each of its frames
- a single Gaussian model for the collected MFCC values is computed. Specifically, we collect its mean, covariance and inverse covariance. Mean is a 13 size vector, while covariance and inverse covariance are 13x13 matrices. If a problem of ill-conditioned covariance matrices is encountered (i.e., a not positive semi-definite covariance matrix has been computed), only values of the diagonal of these problematic covariance matrices are used. This has allowed to avoid the presence of outliers when computing similarity, while still taking into account excerpts for which a covariance matrix of the MFCC values could not be correctly computed.
- based on the HPCP values computed by Echo Nest, we use Essentia’s `Key Detector` to associate a key to each first and fourth beat of the bar. The reason why we keep values for these two particular beats is that this allows us to perform a more precise tonal comparison when trying to merge two excerpts in the real-time application: if the key of the first beat of the inspected excerpt is very different from the key of the fourth beat of the excerpt for which we’re looking for similar pieces, the candidate is discarded.

This procedure is repeated for each excerpt, in order to get a deep description for all of them and perform more precise similarity computation in the real-time application. In addition, we store some additional level-song descriptors, specifically artist, title and year of release, and acousticness (computed with Echo Nest). Finally, for each song we create a corresponding JSON file in which we store all the descriptors computed.

The list of descriptors computed during this task is summarized in table 5.1.

---

<sup>1</sup>Essentia uses the MFCC-FB40 implementation, which decomposes the signal into 40 bands from 0 to 11000Hz, takes the log value of the spectrum energy in each mel band and finally applies a Discrete Cosine Trasform of the 40 bands down to 13 mel coefficients.

Name	Source	Level	Reason
Title, Artist, Year	Provided	Song-Level	Display more information about the current playing track in the GUI
Acousticness	Echo Nest	Song-Level	Give the user the chance to filter music in regards to its nature (acoustic or electronic music)
MFCC	Essentia	Bar-Level	Timbre similarity computation
BPM	Echo Nest	Bar-Level	Avoid consecutive excerpts with very different BPM
Onset Rate	Essentia	Bar-Level	Give the user the chance to filter music in regards to the presence of percussive elements
Dissonance	Essentia	Bar-Level	Give the user the chance to filter music in regards to the dissonance <sup>2</sup> of excerpts
Loudness	Echo Nest	Bar-Level	Give the user the chance to filter music in regards to its loudness
Bark Bands	Essentia	Bar-Level	Give the user the chance to filter music in regards to its “sparseness”, i.e. the amount of mel bands with significant energy level
HPCP	Echo Nest	Beat-Level	Use them to compute key
Key	Essentia	Beat-Level	Use them to discard the possibility of having two consecutive dissonant excerpts in the playlist

TABLE 5.1: Descriptors computed by the offline application.

## 5.2 FastMap computation

The procedure just described for computing descriptors give us a 410 size vector for each excerpt, and a total number of 159239 excerpts.

In order to achieve good real-time performance when comparing these excerpt, a dimensionality reduction of these vectors is required. Furthermore, as seen in 2.3 the computation of Kullback-Leibler divergence, although showing very good results in capturing the timbre similarity, is a very intensive computational operation and therefore a simpler distance measure with comparable results is preferred.

<sup>2</sup>During development, it has been empirically noticed that dissonance has a significant correlation to the perception of noise: the more dissonant an excerpt is, the more it is perceived as noisy.

These requirements were faced also in [51], in which Schnitzer et al. present a filter-and-refine method to speed up nearest neighbor searches with the Kullback-Leibler divergence for multivariate Gaussians, yielding high recall values of 95-99% compared to a standard linear search. The original FastMap was proposed in 1995 by Faloutsos and Lin [57] for indexing and data-mining multimedia datasets. It was used for the first time for computationally heavy, non-metric measures and nearest neighbor retrieval in [55], for speeding up classification of handwritten digits. FastMap was used for the first time in MIR by Cano et al. in [56] in the attempt of reducing high dimensional music timbre similarity space into a 2-dimensional space. This was done not for speeding up classification, but rather for visualization purposes.

The idea behind the use of a FastMap for classification or computing similarities is to compute with the original distance measure  $D()$  (computationally intensive) just a subset of all the distances, specifically the distances between each point and a subset of  $2k$  points (the “*pivots*”); then, on the basis of these distances computed, each feature vector is mapped with a non-linear transformation into a point of a  $k$ -dimension space, where a simpler distance measure can be applied, with a little loss of accuracy.

### Original FastMap

For choosing the pivot elements, the original FastMap [57] follows this strategy:

- $k$  element  $x_1^1, x_2^1, \dots, x_k^1$  are randomly selected from the collection of feature vectors
- for each  $x_i^1$ , we pick its corresponding most distant object  $x_i^2$  using the original distance measure  $D()$

### The filter-and-refine method

Unfortunately, this has shown bad performance

## Chapter 6

# The Real-Time Application

### 6.1 Implementation (python server + html client), Gstreamer

### 6.2 Functioning

Descriptors of first bar, similarity computation (both as an Euclidean Distance and as SKL)

## Part III

# Results and Discussion





## Chapter 7

# Evaluation

Also some words on the Kiosk

## Chapter 8

# Future Work

## Appendix A

# List of Essentia descriptors

As of November 2014, the features provided by Essentia 2.0.1 are:

Category	Name
lowLevel	average loudness
rhythm	bpm
rhythm	onset rate
sfx	pitch after max to before max energy ratio
sfx	pitch centroid
sfx	pitch max to total
sfx	pitch min to total
tonal	chords changes rate
tonal	chords number rate
tonal	key strength
tonal	tuning diatonic strength
tonal	tuning equal tempered deviation
tonal	tuning frequency
tonal	tuning nontempered energy ratio
lowLevel	barkbands kurtosis
lowLevel	barkbands skewness
lowLevel	barkbands spread
lowLevel	dissonance
lowLevel	hfc
lowLevel	pitch
lowLevel	pitch instantaneous confidence
lowLevel	pitch salience

lowLevel	silence rate 20dB
lowLevel	silence rate 30dB
lowLevel	silence rate 60dB
lowLevel	spectral centroid
lowLevel	spectral complexity
lowLevel	spectral crest
lowLevel	spectral decrease
lowLevel	spectral energy
lowLevel	spectral energyband high
lowLevel	spectral energyband low
lowLevel	spectral energyband middle high
lowLevel	spectral energyband middle low
lowLevel	spectral flatness db
lowLevel	spectral flux
lowLevel	spectral kurtosis
lowLevel	spectral rms
lowLevel	spectral rolloff
lowLevel	spectral skewness
lowLevel	spectral spread
lowLevel	spectral strongpeak
lowLevel	zerocrossingrate
rhythm	beats loudness
rhythm	first peak bpm
rhythm	first peak spread
rhythm	first peak weight
rhythm	second peak bpm
rhythm	second peak spread
rhythm	second peak weight
sfx	inharmoniccity
sfx	oddtievenharmonicenergyratio
tonal	chords strength
rhythm	beats position
rhythm	bpm estimates
rhythm	bpm intervals
rhythm	onset times
tonal	chords histogram
tonal	thpcp

lowLevel	barkbands
lowLevel	mfcc
lowLevel	sccoeffs
lowLevel	scvalleys
rhythm	beats loudness band ratio
sfx	tristimulus
tonal	hpcp
tonal	chords key
tonal	chords scale
tonal	key key
tonal	key scale
tonal	chords progression

TABLE A.1: List of features computable with Essentia.

## Appendix B

### List of Echonest Features

Category	Name
----------	------

TABLE B.1: List of audio features provided by Echonest.

## Appendix C

### Phonos: list of songs

The musical pieces to be used during the “*Phonos, 40 anys de música electrònica a Barcelona*” exhibition at Museu de la Musica (L’Auditori, Carrer de Lepant, 150, 08013 Barcelona) are:

<b>Title</b>	<b>Album Artist</b>	<b>Year</b>
Trois Cánones en Hommage à Galilea		1968
De Dos Para Uno	Alain Perón	1996
Los Edictos	Alain Perón	1998
Nexus	Albert Llanas	1999
Formants	Albert Llanas	2004
Monoleg	Alejandro Martínez	
Helesponto	Alejandro Martínez	1982
Tazir	Alejandro Martínez	1984
Crisálida	Alejandro Martínez	1987
Machina animata	Alejandro Martínez	1987
Canción de Otoño	Alejandro Martínez	1989
Homenaje L.Nono	Alejandro Martínez	1990
Música Palimpsesto	Alejandro Martínez	1991
Vaciando el hueco	Alejandro Martínez	1996
Témenos	Alex Arteaga	2006
Panales	Alex Geell	2010
Fluir	Alex Sanjurjo	2003
Ayehli	Alexandra Gardner	2002
Snapdragon	Alexandra Gardner	2002
New Skin	Alexandra Gardner	2003

Onice	Alexandra Gardner	2003
Luminoso	Alexandra Gardner	2004
Tourmaline	Alexandra Gardner	2004
Apparatus, Experimentalis	Alexandre Marino	2008
Apparatus, Musical	Alexandre Marino	2008
Joc - Eventos	Andrés Lewin-Richter	1976
Joc - Fondo	Andrés Lewin-Richter	1976
Acción 2 - 1	Andrés Lewin-Richter	1978
Acción 2 - 2	Andrés Lewin-Richter	1978
Acción 2 - 3	Andrés Lewin-Richter	1978
Acción 2 - 4	Andrés Lewin-Richter	1978
Giravolt	Andrés Lewin-Richter	1978
El Paraiso	Andrés Lewin-Richter	1979
El Viento I - 1	Andrés Lewin-Richter	1979
El Viento I - 2	Andrés Lewin-Richter	1979
El Viento I - 3	Andrés Lewin-Richter	1979
El Viento II	Andrés Lewin-Richter	1979
Reacciones I II	Andrés Lewin-Richter	1979
Secuencia IV	Andrés Lewin-Richter	1979
Baschetiada	Andrés Lewin-Richter	1980
El Viento III	Andrés Lewin-Richter	1980
El Viento IV	Andrés Lewin-Richter	1980
Reacciones III	Andrés Lewin-Richter	1980
Wagler Walricci	Andrés Lewin-Richter	1981
Actualidad discográfica	Andrés Lewin-Richter	1982
Sones	Andrés Lewin-Richter	1982
6 Songs	Andrés Lewin-Richter	1983
Quorum	Andrés Lewin-Richter	1983
Secuencia V	Andrés Lewin-Richter	1983
Secuencia VI	Andrés Lewin-Richter	1983
Tinell	Andrés Lewin-Richter	1983
Cogida	Andrés Lewin-Richter	1984
In memoriam Manuel Valls	Andrés Lewin-Richter	1984
Isaac el Cec	Andrés Lewin-Richter	1984
Juegos	Andrés Lewin-Richter	1985
Musica electroacústica	Andrés Lewin-Richter	1985
Solars Vortices	Andrés Lewin-Richter	1985



Desfigurat	Andrés Lewin-Richter	1986
Diálogos	Andrés Lewin-Richter	1987
Secuencia VII	Andrés Lewin-Richter	1987
Homenaje a Zinovieff	Andrés Lewin-Richter	1988
Secuencia VIII	Andrés Lewin-Richter	1988
Verra la Morte	Andrés Lewin-Richter	1988
Verra la Morte 1	Andrés Lewin-Richter	1988
Verra la Morte 2	Andrés Lewin-Richter	1988
Verra la Morte 3	Andrés Lewin-Richter	1988
Verra la Morte 4	Andrés Lewin-Richter	1988
Verra la Morte 5	Andrés Lewin-Richter	1988
Verra la Morte 6	Andrés Lewin-Richter	1988
Verra la Morte 7	Andrés Lewin-Richter	1988
Verra la Morte 8	Andrés Lewin-Richter	1988
99 Golpes	Andrés Lewin-Richter	1989
Ben avra questa donna cor di ghiacio	Andrés Lewin-Richter	1989
Secuencia IX	Andrés Lewin-Richter	1989
Strings	Andrés Lewin-Richter	1989
Brossiana	Andrés Lewin-Richter	1990
Donne Fiori	Andrés Lewin-Richter	1990
Fragmento (a Nono)	Andrés Lewin-Richter	1990
Frullato	Andrés Lewin-Richter	1990
Ludus Basiliensis	Andrés Lewin-Richter	1991
Reacciones IV	Andrés Lewin-Richter	1991
Secuencia X	Andrés Lewin-Richter	1991
Radio 2	Andrés Lewin-Richter	1996
Sarangi	Andrés Lewin-Richter	1999
Configuraciones	Andrés Lewin-Richter	2000
Constelaciones	Andrés Lewin-Richter	2000
Figuras	Andrés Lewin-Richter	2000
Resonancias	Andrés Lewin-Richter	2000
Secuencia XI	Andrés Lewin-Richter	2001
Secuencia XII	Andrés Lewin-Richter	2001
Dreams	Andrés Lewin-Richter	2002
Ludus Allavarium	Andrés Lewin-Richter	2002
Platjes	Andrés Lewin-Richter	2002
Secuencia XIII	Andrés Lewin-Richter	2002

Signals	Andrés Lewin-Richter	2002
Viso di Primavera	Andrés Lewin-Richter	2002
Fantasia	Andrés Lewin-Richter	2003
Juego de Acordeón	Andrés Lewin-Richter	2003
Meisoh No Ne	Andrés Lewin-Richter	2003
Melodias	Andrés Lewin-Richter	2003
Metálica	Andrés Lewin-Richter	2003
Omaggio a Berio: sequenza per tuba	Andrés Lewin-Richter	2003
Secuencia XIV	Andrés Lewin-Richter	2003
Essay on Trombone	Andrés Lewin-Richter	2004
Fragments	Andrés Lewin-Richter	2004
Secuencia XV	Andrés Lewin-Richter	2004
Arssonxx.rne	Andrés Lewin-Richter	2005
Fluxus es zen?	Andrés Lewin-Richter	2005
Interacciones	Andrés Lewin-Richter	2006
On "Freesound" Water	Andrés Lewin-Richter	2006
Secuencia XVI	Andrés Lewin-Richter	2006
For Harry	Andrés Lewin-Richter	2007
Retales	Andrés Lewin-Richter	2007
Sombras	Andrés Lewin-Richter	2007
Soplos	Andrés Lewin-Richter	2007
Sospiri	Andrés Lewin-Richter	2007
Friendship Quartet	Andrés Lewin-Richter	2008
Homenaje a Pierre Schaeffer	Andrés Lewin-Richter	2008
Makeup	Andrés Lewin-Richter	2008
Schaeffer granulado	Andrés Lewin-Richter	2008
Aire	Andrés Lewin-Richter	2009
Génesis	Andrés Lewin-Richter	2009
Homenaje a Varese	Andrés Lewin-Richter	2009
Memento	Andrés Lewin-Richter	2009
Paseo BCN	Andrés Lewin-Richter	2009
Sancta Maria	Andrés Lewin-Richter	2009
Slapring	Andrés Lewin-Richter	2009
Spring	Andrés Lewin-Richter	2009
Imagenes	Andrés Lewin-Richter	2010
Secuencia XVIII Fagot	Andrés Lewin-Richter	2010
Multifonia	Andrés Lewin-Richter	2011

Campanas para una celebracion	Andrés Lewin-Richter	2012
Multifonia III	Andrés Lewin-Richter	2012
Secuencia XIX	Andrés Lewin-Richter	2014
Espai Sonor	Anna Bofill	
Trio para Violin y Cinta	Anna Bofill	
Sinapsis	Ariadna Alsina	2006
Reconstrucció	Ariadna Alsina	2011
Vels Vitris	Ariadna Alsina	2012
Contramarea	Ariadna Alsina & David Salleras	2009
La Música Que Había en Mis Objetos	Arturo Moya	1996
Estampas de Caza 1	Arturo Moya	2000
Estampas de Caza 2	Arturo Moya	2000
Estampas de Caza 4	Arturo Moya	2000
Estampas de Caza 5	Arturo Moya	2000
Estate quieto Voltaire	Arturo Palaudarias	
Adolescencia y Estrella	Arturo Palaudarias	1980
Escudellers	Arturo Palaudarias	1981
Piamo	Arturo Palaudarias	1984
Toda la Memoria de un Hombre	Arturo Palaudarias	1987
El Destino de las Cosas	Arturo Palaudarias	1988
La Luz de los Sueños	Arturo Palaudarias	1989
Boule de Feu	Arturo Palaudarias	1990
Paréntesis militar	Arturo Palaudarias	1990
El Juicio Estético Universal	Arturo Palaudarias	1991
Moverse en el Tiempo	Arturo Palaudarias	1997
Women in Process	Aurelio Edler-Copes	2013
Exquisits	Cadavers	2003
Latido	Carlos Lupprián	1995
Agugagá	Carlos Lupprián	1996
Naturaleza Muerta	Carlos Lupprián	1997
Improvisación con Oratio Trio	Claudio Nervi	2010
Valent La Notte	Claudio Zulian	
El Libro de los Excesos	Claudio Zulian	1983
San Claudio Vive Solo	Claudio Zulian	1985
Sexo y Política	Claudio Zulian	1987
I Quattro Continenti	Claudio Zulian	1989

Por de Ser Set	Claudio Zulian	1989
Sueños Ecléctricos	Claudio Zulian	1989
Variazione Angelica	Claudio Zulian	1990
2 Escenas de Macbeth - 1	Claudio Zulian	1991
2 Escenas de Macbeth - 2 Ruidos	Claudio Zulian	1991
Armonias 1	Concha Trallero	1980
Armonias 2	Concha Trallero	1980
Armonías Sonoras 1	Concha Trallero	1980
Armonías Sonoras 2	Concha Trallero	1980
Leftrarur, Viajero Ensoñado - El Río de la Vida	Cristián López	2005
Leftrarur, Viajero Ensoñado - Espíritu Azul	Cristián López	2005
Leftrarur, Viajero Ensoñado - Interludio	Cristián López	2005
Leftrarur, Viajero Ensoñado - Piedra Solitaria	Cristián López	2005
Leftrarur, Viajero Ensoñado - Relámpago Azul	Cristián López	2005
Relief II	Cristián Morales-Ossio	2001
TRTPS	Daniel Domínguez Teruel	2010
SKTHN	Daniel Domínguez Teruel	2012
Study I	Daniel Domínguez Teruel	2013
Study II	Daniel Domínguez Teruel	2013
Study V	Daniel Domínguez Teruel	2013
Say It	Daniel Rios Aranda	1987
Erial	Daniel Rios Aranda	1990
Sueños	Danilo Vidotti	2008
Psicofonias Urbanas 1	Danio Catanuto	2010
Psicofonias Urbanas 2	Danio Catanuto	2010
Formantes	Darío Cortés	1998
Pulsajes	David Dalmazzo	2010
Confluencies	David Padros	1985
Caosmofonia	Diego Dall'Osto	1998
Kinoko Tabí	Doénado, el Ur	1988
Pedicoj en la Arena del Pamir	Doénado, el Ur	1989
Zalody	Doénado, el Ur	1990
Yñé do zalod	Doénado, el Ur	1991
A Sensu Contrario	Doénado, el Ur	1992
Blordt Prelar	Doénado, el Ur	1992
Kzadzak	Doénado, el Ur	1994
Tu Mateix	Edgar Barroso	2004

Dux	Edgar Barroso	2005
Tau	Edgar Barroso	2005
Tu Soplo Que Transporta	Edgar Barroso	2005
IOD	Edgar Barroso	2006
CYT	Edgar Barroso	2007
Mármore	Edson Zampronha	2001
Mármore 1	Edson Zampronha	2001
Mármore 2	Edson Zampronha	2001
Mármore 3	Edson Zampronha	2001
Read my LISP	Eduard Resina	1991
L'Esquizofrènia Dels Sons	Eduard Resina	1993
Aca Amaron	Eduard Resina	2001
L'Anna-crusa	Eduard Resina	2002
Espai Sonor	Eduardo Polonio	1976
Requiem per una Veu Perduda	Eduardo Reck Miranda	1997
Midi de Sable	Elsa Justel	2000
Elementos Constantes, Hechos Variables	Enrique Marín	2002
Transiciones de Fase	Enrique Marín	2007
Untitled 1	Ensamble Crumble y Re- acTable	2006
Untitled 2	Ensamble Crumble y Re- acTable	2006
Untitled 3	Ensamble Crumble y Re- acTable	2006
Untitled 4	Ensamble Crumble y Re- acTable	2006
Untitled 5	Ensamble Crumble y Re- acTable	2006
Untitled 6	Ensamble Crumble y Re- acTable	2006
Untitled 7	Ensamble Crumble y Re- acTable	2006
Untitled 1	FMOL Trio	2001
Untitled 2	FMOL Trio	2001
Untitled 3	FMOL Trio	2001
CampoSanto	Felipe Pérez Santiago	2004
Encandilado	Felipe Pérez Santiago	2007

Hunger FM	Felipe Pérez Santiago	2009
Hurt	Felipe Pérez Santiago	2009
Ishmael	Felipe Pérez Santiago	2009
Miuk	Felipe Pérez Santiago	2009
Post War	Felipe Pérez Santiago	2009
Tacto	Felipe Pérez Santiago	2009
War-Post War	Felipe Pérez Santiago	2009
Pronto Desapareceremos	Felipe Pérez Santiago	2012
Ecos 1	Fernando Jobke	2008
Ecos 2	Fernando Jobke	2008
Ecos 3	Fernando Jobke	2008
Ecos 4	Fernando Jobke	2008
Ecos 5	Fernando Jobke	2008
Ecos 6	Fernando Jobke	2008
Cuerpos Sensibles	Félix Luque & Ricardo Gadea	2005
The Machine Manifesto	Félix Luque & Thomas Charveriat	2004
Batucada Amenazante	Gabriel Brncic	1970
El Túnel (a Ernesto Sabato)	Gabriel Brncic	1970
Agua 1	Gabriel Brncic	1971
Agua 2	Gabriel Brncic	1971
Agua 3	Gabriel Brncic	1971
Cielo	Gabriel Brncic	1980
Destierro	Gabriel Brncic	1980
Chile Fértil Provincia	Gabriel Brncic	1983
Concert Gothique	Gabriel Brncic	1985
Operas Rotas	Gabriel Brncic	1985
Clarinen Tres	Gabriel Brncic	1986
Clarinen Tres	Gabriel Brncic	1986
Desêtre a Oscar Masotta	Gabriel Brncic	1986
Triunfo Para las Madres	Gabriel Brncic	1986
Aria y Pasacalle	Gabriel Brncic	1987
Ese Mar	Gabriel Brncic	1987
Música de cámara	Gabriel Brncic	1987
Historia de Dos Ciudades	Gabriel Brncic	1988
Alegrias	Gabriel Brncic	1989
Composición de 1989 a Eduardo Polonio	Gabriel Brncic	1989

Dulcian Concert	Gabriel Brncic	1989
ariaciones sobre Sonatas e Interludios	Gabriel Brncic	1989
Adagio-Scherzo	Gabriel Brncic	1990
Vade Retro a Luigi Nono	Gabriel Brncic	1990
Dos Esbozos Para Antiguos Instrumentos Electrónicos	Gabriel Brncic	1994
...Que No Desorganitza Cap Murmuri	Gabriel Brncic	1995
Constanza	Gabriel Brncic	1996
Claro-Oscuro	Gabriel Brncic	1998
Meng	Gabriel Brncic	1998
Clarinet Concert	Gabriel Brncic	1999
Coreutica	Gabriel Brncic	1999
Ergon-Rondeau	Gabriel Brncic	2000
A Joan Miró	Gabriel Brncic	2001
Alto-Concert II	Gabriel Brncic	2001
Bass clarinet-Concert for Harry Sparnaay	Gabriel Brncic	2003
Son(ru)idos I	Gabriel Brncic	2003
Son(ru)idos II	Gabriel Brncic	2003
La Casa del Viento 1	Gabriel Brncic	2006
La Casa del Viento 2	Gabriel Brncic	2006
Pregoneros de Barcelona	Gaspar Lukacs Esguep	2002
Sin título	Germán Brull Moreno	2004
Sin título	Germán Brull Moreno	2004
Arboleda	Graciela Muñoz Farida	2011
Fragments de un Arbol	Graciela Muñoz Farida	2011
Lo Que No Das Te Lo Quitas	Graciela Muñoz Farida	2011
Viento Sur	Graciela Muñoz Farida	2011
Piece for Guitar and Tape	Graeme Truslove	2001
Improvisation	Graham Coleman	2007
Improvisation	Graham Coleman	2007
Guitarrísticamente	Guillermo Eisner	2007
Duo Para Siete	Igor Bimsbergen	1996
Luis y Marylin	Igor Bimsbergen	1998
Free What	Ismael Sanoja & Kai Kraatz	2006
Free What 1	Ismael Sanoja & Kai Kraatz	2006
Free What 2	Ismael Sanoja & Kai Kraatz	2006
Free What 3	Ismael Sanoja & Kai Kraatz	2006

Free What 4	Ismael Sanoja & Kai Kraatz	2006
Free What 5	Ismael Sanoja & Kai Kraatz	2006
Free What 6	Ismael Sanoja & Kai Kraatz	2006
Free What 7	Ismael Sanoja & Kai Kraatz	2006
Free What 8	Ismael Sanoja & Kai Kraatz	2006
Traumtantze	Jan Schacher	2000
Preludios	Javier Navarrete	1976
Almogavers	Jelena Vico	2008
Brithm	Jelena Vico	2008
Mrzbw	Jelena Vico	2008
Pangea	Jelena Vico	2008
Zeno	Jelena Vico	2008
Zitar	Jelena Vico	2008
Gallinaria	Jep Nuix	1980
Doble Pea de Lletres i Sons	Jep Nuix	1981
Tres Canons de Noces	Jep Nuix	1981
Ad Valorem	Jep Nuix	1984
Halterofilia 1	Jep Nuix	1984
Serenata Nocturna	Jep Nuix	1985
L’Inizio	Jep Nuix	1986
Dit a Dit, Pas a Pas	Jep Nuix	1988
Asirara	Jep Nuix	1989
Monoleg	Jep Nuix	1989
Trialeg	Jep Nuix	1989
His Master’s Voice	Jep Nuix	1990
Improvisacio per a tubs	Jep Nuix	1990
Pensant en Nono	Jep Nuix	1990
Percuflu	Jep Nuix	1990
Atentament	Jep Nuix	1992
Stack	Jep Nuix	1995
Intersections-BouleWav 2.0	Joan Bages i Rubı	2006
Al Tranquilodromo	Joan Josep Ordinas & Claudio Zulian	1981
Passadis	Joan Sanmarti	1200
Reflexos Improvisaciones Asistidas por Ordenador	Joan Sanmarti	1997
Xtrapolucio 4	Joan Sanmartı	1998



Ricercare a 5	Jordi Rossinyol	1986
Objectes Trobats a la Platja	Jordi Rossinyol	1987
Ocellots	Jordi Rossinyol	1988
Mòbils Inquiets i Altres Equívocs	Jordi Rossinyol	1989
Prosper Laberint Intermitent	Jordi Rossinyol	1990
Variaciones guit	Jordi Rossinyol	1990
Concert Mestis	Jordi Rossinyol	1997
Ecliptic	Jordi Rossinyol	2004
El Doble Bandoneón	Jorge Sad	1998
La Ida Hacia Abajo de la Tierra de la Tarde	Jorge Sad	1999
Landscape	Josep Maria Guix	2010
Landscape	Josep Maria Guix	2010
Landscape	Josep Maria Guix	2010
Oxo	Josep Maria Mestres Quadreny	1963
Peça per a Serra Mecanica	Josep Maria Mestres Quadreny	1963
Homenaje a Galileo	Josep Maria Mestres Quadreny	1965
Aronada	Josep Maria Mestres Quadreny	1972
El Teler de Teresa Codina	Josep Maria Mestres Quadreny	1973
Song for Jane Manning	Josep Maria Mestres Quadreny	1973
Espai Sonor	Josep Maria Mestres Quadreny	1976
Espai Sonor	Josep Maria Mestres Quadreny	1976
Quina	Josep Maria Mestres Quadreny	1979
Cánones a Galileo	Josep Maria Mestres Quadreny	1989
El Pensamiento Que Se Trabaja Hacia la Luz	José Manuel Berenguer	
Spira	José Manuel Berenguer	
Montardo	José Manuel Berenguer	1983
A Florats	José Manuel Berenguer	1984
La Logica de la Sorpresa	José Manuel Berenguer	1984
El Ponent Excesiu	José Manuel Berenguer	1985
La Perla Estranya	José Manuel Berenguer	1985
La Relojeria del Tío Paco	José Manuel Berenguer	1985
Música en la Noche	José Manuel Berenguer	1985
Quartet Ambar	José Manuel Berenguer	1986
Color	José Manuel Berenguer	1987
Polifonía de Colores	Juan Antonio Moreno	1984
Preludio III a Lluís Callejo	Juan Antonio Moreno	1988
Nono Está Aquí	Juan Antonio Moreno	1990

Buenhache	Juan Antonio Moreno	1991
G-Gems	Lina Bautista	
Bombyx Mori	Lina Bautista	2010
Encélado	Lina Bautista	2011
A River From the Walls	Linda Antas	1999
Sueño sin palabras	Linda Antas	2001
Untitled	Lisos-Estriados	2001
Carota i Caramel	Llorenç Balsach	1976
Espais residuals (Espai I)	Llorenç Balsach	1976
L'assassi Bagliatti	Llorenç Balsach	1977
El Cant de les Arteries	Llorenç Balsach	1979
Caleidoscopi	Lluis Callejo	
Dibuixos	Lluis Callejo	1981
Estructures 6502	Lluis Callejo	1982
Paisatges	Lluis Callejo	1983
Tèxtils	Lluis Callejo	1984
A Pitàgores en do	Lluis Callejo	1985
A Pitàgores en re	Lluis Callejo	1985
Espai Sonor	Lluis Callejo	2003
Stokos IV	Lluis Callejo	2003
La Triste Herida de Margot	Luis Caruana	2001
Por Tus Pliegues Transita la Pena	Luis Caruana	2001
Animales Divinos	Marcelo DeMatei & Carlos Smith	2003
Petit Estudi	Mario Peña y Lillo	
Beso	Mario Peña y Lillo	2013
El Contorno de sus Ojos	Mario Peña y Lillo	2013
Esencia	Mario Peña y Lillo	2013
He Perdido la Apuesta	Mario Peña y Lillo	2013
Youkali	Mario Peña y Lillo	2013
Figuras Negras	Mario Verandi	1992
Flamencas	Mario Verandi	1995
Faces and Intensities	Mario Verandi	1996
Frèquences de Barcelone	Mario Verandi	1997
Mu	Mario Verandi	1997
Mists	Matthew Burtner	1996
Fern	Matthew Burtner	1997

Incantation S4	Matthew Burtner	1997
Split Voices	Matthew Burtner	1997
Glass Phase	Matthew Burtner	1998
Portals of Distortion	Matthew Burtner	1998
Delta 1	Matthew Burtner	2000
Duo	Mauricio Valdés	2002
Popan II	Mauricio Valdés	2008
Gramatges	Mercè Capdevila	1983
Baobab	Mercè Capdevila	1985
Nu	Mercè Capdevila	1990
Alegries de Comèdia	Mercè Capdevila	1991
Mercuri	Mercè Capdevila	1991
Fons de Mar	Mercè Capdevila	2000
Pols	Mercè Capdevila	2000
Puente	Mercè Capdevila	2000
A Chillida	Mercè Capdevila	2009
Time Machine	Miquel Jordà	2000
La Máquina, el Humano y el Olivo	Nadine Kroher	2013
Mixed Signals	Nadine Kroher	2014
Concierto Sonocromático	Neil Harbisson	2011
Catarsis III	Oliver Rappoport	2009
Laberint Mutant II	Oriol Graus	1987
Miradaclosa IV	Oriol Graus	1987
I despres...	Oriol Graus	1990
La Solitud de l'Origen	Oriol Graus	1990
La conseqüència	Oriol Graus	1990
La intuïció	Oriol Graus	1990
Oketus	Oriol Graus	1990
Diferents Formes de Dir - T'Ho	Oriol Graus	1991
La Tolerancia	Oriol Graus	1993
El Laberint de l'Esperança	Oriol Graus	2000
Paisatge Interior	Oriol Graus	2010
Black Nature	Oscar Martin	2012
Black Nature	Oscar Martin	2012
Fer et Defer	Pablo Fredes	
Historia del Vinilo	Pablo Fredes	
Trama	Pablo Fredes	

Las Nenas del Sonido	Pablo Fredes	2002
Ça Fait Faire Ça Ruidos	Pablo Fredes	2004
El Círculo de Cero	Pablo Fredes	2009
sX-off-on	Pablo Fredes	2009
Azu Gemma Torralbo	Pablo Fredes	2011
Son-ethos (Sueños en el Sueño)	Pablo Fredes	2011
Son-file	Pablo Fredes	2011
iO	Pablo Fredes	2011
on_off Gemma Torralbo	Pablo Fredes	2011
Cero Roce Sostenuto	Pablo Fredes	2012
Estratos	Pedro Barboza	2001
Estratos	Pedro Barboza	2001
La fila de Ocata	Pedro Barboza	2001
inTENSIONtres	Pedro Barboza	2004
Mantra I	Ramon Humet	2005
1	Rebecka Biro	2005
2	Rebecka Biro	2005
Daffodil for Peter Billings	Ricardo Arias	
Improvisación	Ricardo Arias & Carlos Gómez	2009
Sol Sonoro 1	Ricardo Arias & Roberto García	2008
Sol Sonoro 2	Ricardo Arias & Roberto García	2008
Je Suis l'Autre	Roger Costa	2012
off ICMC2005	Ross Bencina	2005
off ICMC2005	Ross Bencina	2005
Simple Math	Sanjay Fernandes	2010
Ella Era Todo - Escribir Sobre Piel	Sebastián García Ferro	
Ella Era Todo - Yang	Sebastián García Ferro	
Europa 1 - Piano	Sebastián García Ferro	
Europa 2 - Crescendo	Sebastián García Ferro	
Europa 3 - Bosque	Sebastián García Ferro	
Europa 4 - Vibracion	Sebastián García Ferro	
Europa 5 - Noise Delay Long	Sebastián García Ferro	
Europa 6 - Piano	Sebastián García Ferro	
Equus	Sebastián García Ferro	2001
Noise	Sebastián García Ferro	2001

Pulso	Sebastián García Ferro	2001
Afro Dero	Sebastián García Ferro	2002
Ceratti	Sebastián García Ferro	2002
Dash	Sebastián García Ferro	2002
Seed	Sebastián García Ferro	2002
Shadow	Sebastián García Ferro	2002
Silla	Sebastián García Ferro	2002
Absorción Vertical	Sebastián García Ferro	2003
Bosa	Sebastián García Ferro	2003
Drugs	Sebastián García Ferro	2003
Etheric	Sebastián García Ferro	2003
Fiesta	Sebastián García Ferro	2003
Final	Sebastián García Ferro	2003
Huellas	Sebastián García Ferro	2003
Huellas Intro	Sebastián García Ferro	2003
Mistrius	Sebastián García Ferro	2003
Nervio	Sebastián García Ferro	2003
Rebotes	Sebastián García Ferro	2003
Rhesus	Sebastián García Ferro	2003
Sentadas	Sebastián García Ferro	2003
Solo Caro	Sebastián García Ferro	2003
Trio	Sebastián García Ferro	2003
Viaje Transparente	Sebastián García Ferro	2003
Vacio y Multitud 1	Sebastián García Ferro	2004
Vacio y Multitud 2	Sebastián García Ferro	2004
Bajo el Agua	Sebastián García Ferro	2005
Caidas	Sebastián García Ferro	2005
Come Home	Sebastián García Ferro	2005
Flotar	Sebastián García Ferro	2005
Sumergir	Sebastián García Ferro	2005
Back (escena 1)	Sebastián García Ferro	2006
Back (escena 3)	Sebastián García Ferro	2006
Back (escena 5 y 6)	Sebastián García Ferro	2006
Gatos	Sebastián García Ferro	2006
Mandrös	Sebastián García Ferro	2006
Modified - Intro	Sebastián García Ferro	2006
Peces	Sebastián García Ferro	2006

Caras Jazzie End	Sebastián García Ferro	2007
Clock	Sebastián García Ferro	2007
Corn	Sebastián García Ferro	2007
Despertar	Sebastián García Ferro	2007
Fork	Sebastián García Ferro	2007
Mañana	Sebastián García Ferro	2007
Mediodia	Sebastián García Ferro	2007
Metting	Sebastián García Ferro	2007
Noche	Sebastián García Ferro	2007
Pointing	Sebastián García Ferro	2007
Sueños	Sebastián García Ferro	2007
Tarde	Sebastián García Ferro	2007
Vaiven Parte 1	Sebastián García Ferro	2007
Vaiven Parte 2	Sebastián García Ferro	2007
Travellers 1	Sebastián García Ferro	2008
Travellers 2	Sebastián García Ferro	2008
Travellers 3	Sebastián García Ferro	2008
La Lámpara	Sebastián Jara Bunster	2010
For Eric	Sergi Jordá	2001
Big Bang	Sergio Naddei	2011
Rock Memories	Sergio Naddei	2011
The Fly	Sergio Naddei	2011
Windows	Sergio Naddei	2012
Almost New Places	Sergio Naddei	2013
Almost New Spaces	Sergio Naddei	2013
Through Memories 1	Sergio Naddei	2013
Through Memories 2	Sergio Naddei	2013
Through Memories 3	Sergio Naddei	2013
Through Memories 4	Sergio Naddei	2013
Through Memories 5	Sergio Naddei	2013
Reactable	Sergio Naddei	2014
Actions	Sergio Poblete	1998
Místicos I Phonos Fund.Miro	Sáez,Ignacio	1987
El Riu Fosc	Sáez,Ignacio	1988
Horizonte Encadenado	Sáez,Ignacio	1990
For Fernando Riera	Teruyoshi Kamiya	1996
Dance of Stone	Teruyoshi Kamiya	1998

The Machine Manifesto	Thomas Charveriat & Félix Luque	2004
Hemispherical Glitch Study	Tim Schmele	2013
Neurospaces	Tim Schmele	2013
Waiting	Tim Schmele	2013
Seguiriyas	Trino Zurita & Teresa Carrasco	2013
Doll_sa.caustika	Xavi Manzanares	2006
Errorunnel	Xavi Manzanares	2006
H2O	Xavi Manzanares	2006
Massiva	Xavi Manzanares	2006
Nnervits	Xavi Manzanares	2006
Nuvols	Xavi Manzanares	2006
Openspaceinvaders	Xavi Manzanares	2006
Plastiknazzxs	Xavi Manzanares	2006
R4gg4gg4r	Xavi Manzanares	2006
Rezzaka	Xavi Manzanares	2006
Segmentationfault0100	Xavi Manzanares	2006
Segmentationfault1001a	Xavi Manzanares	2006
Segmentationfault1001b	Xavi Manzanares	2006
Standbykut	Xavi Manzanares	2006
Stirofoammentre	Xavi Manzanares	2006
Tripikx	Xavi Manzanares	2006
East Cocker	Xavier Maristany	1984
Remember Me	Xavier Maristany	1999

TABLE C.1: Phonos catalogue to be used during the exhibition “*Phonos, 40 anys de música electrònica a Barcelona*”.

# Bibliography

- [1] A. Li-chun Wang, and Th Floor Block F. *An industrial-strength audio search algorithm*. Proceedings of the 4 th International Conference on Music Information Retrieval, 2003.
- [2] J.S. Downie. *The Scientific Evaluation of Music Information Retrieval Systems: Foundations and Future*. Computer Music Journal, 28:12-23, 2004.
- [3] N. Orio. *Music Retrieval: A Tutorial and Review*. Foundations and Trends®in Information Retrieval, 1(1):1-90, 2006.
- [4] M. Schedl, E. Gómez, and J. Urbano. *Music Information Retrieval: Recent Developments and Applications*. Foundations and Trends®in Information Retrieval, 8(2-3):127-261, 2014.
- [5] J.J. Aucouturier, and F. Pachet. *Improving Timbre Similarity: How High is the Sky?*. Journal of Negative Results in Speech and Audio Sciences, 1(1):1-13, 2004.
- [6] X. Serra, M. Magas, E. Benetos, M. Chudy, S. Dixon, A. Flexer, E. Gómez, F. Gouyon, P. Herrera, S. Jorda, O. Paytuvi, G. Peeters, J. Schlüter, H. Vinet, and G. Widmer. *Roadmap for Music Information ReSearch*. Geoffroy Peeters (editor), Creative Commons BY NC ND 3.0 license, 2013.
- [7] M.E.P. Davies, P. Hamel, K. Yoshii and M. Goto. *AutoMashUpper: Automatic Creation of Multi-Song Music Mashups*. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(12):1726-1737, 2014.
- [8] G. Szymanski. *Pandora, or, a never-ending box of musical delights*. Music Reference Services Quarterly, 12(1):21-22, 2009.
- [9] M. Slaney. *Web-scale multimedia analysis: Does content matter?*. IEEE Multimedia, 18(2):12-15, 2011.
- [10] J.J. Aucouturier. *Sounds like teen spirit: Computational insights into the grounding of everyday musical terms*. Language, Evolution and the Brain. Frontiers in Linguistics, 35-64, 2009.



- [11] S.J. Green, P. Lamere, J. Alexander, F. Maillet, S. Kirk, J. Holt, J. Bourque, and X.W. Mak. *Generating transparent, steerable recommendations from textual descriptions of items*. ACM Conference on Recommender Systems (RecSys'09), 281-284, 2009.
- [12] D. Bogdanov, and X. Serra. *From music similarity to music recommendation: Computational approaches based on audio features and metadata*. PhD Thesis, Universitat Pompeu Fabra, 2013.
- [13] D. Schnitzer. *Mirage – High-Performance Music Similarity Computation and Automatic Playlist Generation*. Master's Thesis, Vienna University of Technology, 2007.
- [14] L. Barrington, D. Turnbull, D. Torres, and G. Lanckriet. *Semantic similarity for music retrieval*. Music Information Retrieval Evaluation Exchange (MIREX), 2007.
- [15] K. West, and P. Lamere. *A model-based approach to constructing music similarity functions*. EURASIP Journal on Advances in Signal Processing, 149-149, 2007.
- [16] G. Bonnin, and D. Jannach. *Automated Generation of Music Playlists: Survey and Experiments*. ACM Computing Surveys, 47(2), Article 26, 2014.
- [17] R. Dias, and M. J. Fonseca. *MuVis: An application for interactive exploration of large music collections*. Proceedings of MM, 1043-1046, 2010.
- [18] J.S. Downie. *Music information retrieval*. Annual Review of Information Science and Technology, 37(1):295-340, 2003.
- [19] M. Grachten, M. Schedl, T. Pohle, and G. Widmer. *The ISMIR cloud: A decade of ISMIR conferences at your fingertips*. Proceedings of ISMIR, 63-68, 2009.
- [20] M. Dopler, M. Schedl, T. Pohle, and P. Knees. *Accessing music collections via representative cluster prototypes in a hierarchical organization scheme*. Proceedings of ISMIR, 179-184, 2009.
- [21] I. Schmädecke, and H. Blume. *High performance hardware architectures for automated music classification*. Algorithms from and for Nature and Life, 539-547, 2013.
- [22] O. Celma, and X. Serra. *FOAFing the music: Bridging the semantic gap in music recommendation*. Web Semantics: Science, Services and Agents on the World Wide Web, 6(4):250-256, 2008.
- [23] F. Pachet, P. Roy, and D. Cazaly. *A combinatorial approach to content-based music selection*. Multimedia, 7(1):44-51, 2000.

- [24] D. Bogdanov, and P. Herrera. *How much metadata do we need in music recommendation? A subjective evaluation using preference sets*. WProceedings of ISMIR, 97-102, 2011.
- [25] N. Aizenberg, Y. Koren, and O. Somekh. *Build your own music recommender by modeling internet radio streams*. Proceedings of WWW, 1-10, 2012.
- [26] R. Van Gulik, and F. Vignoli. *Visual playlist generation on the artist map*. Proceedings of ISMIR, 520-523, 2005.
- [27] F. Coelho, J. Devezas, and C. Ribeiro. *Large-scale crossmedia retrieval for playlist generation and song discovery*. Proceedings of OAIR, 61-64, 2013.
- [28] O. Celma. *Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Springer, 2010.
- [29] J. McDermott. *Auditory preferences and aesthetics: Music, voices, and everyday sounds*. Neuroscience of Performance and Choice. Elsevier, 227-256, 2012.
- [30] A. Germain, and J. Chakareski. *Spotify me: Facebook-assisted automatic playlist generation*. Proceedings of MMSP, 25-28, 2013.
- [31] D. Hauger, J. Kepler, M. Schedl, A. Kosir, and M. Tkalcic. *The million musical tweets dataset: What can we learn from microblogs*. Proceedings of ISMIR, 189-194, 2013.
- [32] B. Logan. *Music recommendation from song sets*. Proceedings of ISMIR, 425-428, 2004.
- [33] F. Vignoli, and S. Pauws. *A music retrieval system based on user driven similarity and its evaluation*. Proceedings of ISMIR, 272-279, 2005.
- [34] P. Knees, T. Pohle, M. Schedl, and G. Widmer. *Combining audio-based similarity with Web-based data to accelerate automatic music playlist generation*. Proceedings of MIR, 147-154, 2006.
- [35] D. Jannach, M. Zanker, M. Ge, and M. Gröning. *Recommender systems in computer science and information systems — A landscape of research*. Proceedings of EC-Web, 76-87, 2012.
- [36] N.H. Liu, S.W. Lai, C.Y. Chen, and S.J. Hsieh. *Adaptive music recommendation based on user behavior in time slot*. Computer Science and Network Security, 9(2):219-227, 2009.

- [37] S. McNee, J. Riedl, and J. Konstan. *Being accurate is not enough: how accuracy metrics have hurt recommender systems*. CHI'06 extended abstracts on Human Factors in Computing Systems, p.1001, 2006.
- [38] G. Jawaheer, M. Szomszor, and P. Kostkova. *Comparison of implicit and explicit feedback from an online music recommendation system*. Int. Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec'10), p.47-51, 2010.
- [39] B. Whitman, and S. Lawrence. *Inferring Descriptions and Similarity for Music from Community Metadata*. Proceedings of the 2002 International Computer Music Conference (ICMC 2002), p.591-598, 2012.
- [40] S. Dixon *Onset Detection Revised*. Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx'06), p.133-137, 2006.
- [41] J.P. Bello, C. Duxbury, M. Davies, and M. Sandler. *On the use of phase and energy for musical onset detection in the complex domain*. Signal Processing Letters, IEEE, 11(6):553-556, 2004.
- [42] P. Masri. *Computer Modeling of Sound for Transformation and Synthesis of Musical Signal*. Ph.D. dissertation, Univ. of Bristol, 1996
- [43] J.P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler. *A tutorial on onset detection in music signals*. IEEE Transactions on Speech and Audio Processing, 13(5):1035-1047, 2005.
- [44] S. Böck, and G. Widmer. *Maximum filter vibrato suppression for onset detection*. Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13), Maynooth, Ireland, 2013.
- [45] N. Degara, E.A. Rua, A. Pena, S. Torres-Guijarro, M.E. Davies, and M.D. Plumbley. *Reliability-informed beat tracking of musical signals*. IEEE Transactions on Audio, Speech, and Language Processing, 20(1):290-301, 2012.
- [46] T. Fujishima. *Realtime chord recognition of musical sound: A system using Common Lisp Music*. Proceedings of the International Computer Music Conference, Beijing. 1999.
- [47] E. Gómez. *Tonal Description of Polyphonic Audio for Music Content Processing*. INFORMS Journal on Computing, 18(3):294-304, 2006.
- [48] J. Bonada. *Automatic technique in frequency domain for near-lossless time-scale modification of audio*. Proceedings of International Computer Music Conference, 396-399, 2000.

- [49] G. Cabral, J.P. Briot, and F. Pachet. *Impact of distance in pitch class profile computation*. Proceedings of the Brazilian Symposium on Computer Music, 319-324, 2005.
- [50] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra. *ESSENTIA: an open-source library for sound and music analysis*. Proceedings of the 21st ACM international conference on Multimedia, 855-858, 2013.
- [51] D. Schnitzer, A. Flexer, and G. Widmer. *A fast audio similarity retrieval method for millions of music tracks*. Multimedia Tools and Applications, 58(1):23-40, 2012.
- [52] P. Cano, M. Kaltenbrunner, and N. Wack. *An industrial-strength content-based music recommendation system*. Proceedings of the 28th annual international ACM SIGIR conference on research and development in information retrieval, 673-673, 2005.
- [53] P. Roy, J.J. Aucouturier, F. Pachet, and A. Beurivé. *Exploiting the Tradeoff Between Precision and Cpu-Time to Speed Up Nearest Neighbor Search*. ISMIR, 230-237, 2005.
- [54] M. Levy, and M. Sandler. *Lightweight measures for timbral similarity of musical audio*. Proceedings of the 1st ACM workshop on Audio and music computing multimedia, 27-36, 2006.
- [55] V. Athitsos, J. Alon, S. Sclaroff, and G. Kollios. *Boostmap: A method for efficient approximate similarity rankings*. Proceedings of the 2004 IEEE Computer Society Conference, 2:II-268, 2004.
- [56] P. Cano, M. Kaltenbrunner, F. Gouyon, and E. Batlle. *On the use of FastMap for Audio Retrieval and Browsing*. ISMIR, 2002.
- [57] C. Faloutsos, and K.I. Lin. *FastMap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets*. Proceedings of the 1995 ACM SIGMOD international conference on management of data, 24(2):163-174, 1995.