

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

Abstract	ii
Acknowledgements	iii
Contents	iv
List of Figures	vi
List of Tables	vii
Abbreviations	viii
1 Introduction	1
1.1 The importance of music analysis	1
1.2 Phonos Project	3
1.3 GiantSteps	3
1.4 Purpose of this work	4
1.4.1 Structure of the dissertation	5
I Background	7
2 Music Analysis Techniques	9
2.1 Metadata	9
2.1.1 Computing Similarity With a Vector Space Model	11
2.1.2 Computing Similarity With Co-Occurrence Analysis	11
2.2 Audio Content Analysis	11
2.3 Conceptual Differences Between the Two Approaches	11
3 Computing Music Similarity with Audio Content Descriptors	12
3.1 Literature Review	12
II Methodology	13
4 Phonos Catalogue of Songs	15

5	Computation of Audio Features	16
5.1	Tools used for feature extraction, features extracted	16
5.2	Similarity computation (fast map)	16
6	The Real-Time Application	17
6.1	Implementation (python server + html client), Gstreamer	17
6.2	Functioning	17
III	Results and Discussion	18
7	Evaluation	20
8	Future Work	21
A	List of Essentia Features	22
B	List of Echonest Features	23
C	Phonos: list of songs	24
	Bibliography	25

List of Figures

1.1	Phonos	3
1.2	GiantSteps	5

List of Tables

Abbreviations

LAH List Abbreviations **Here**

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 ¹
- 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 Jukebox², a web-application built upon Echonest 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 Autocanonizer³ and Wub Machine⁴. 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 Tool⁵, 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

¹There 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.

²<http://infinitejuke.com>

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

⁴<http://thewubmachine.com>

⁵http://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 mashup⁶ while also allowing the user a control over the music generated [7]. WRITE MORE ABOUT AUTOMASH HERE

1.2 Phonos Project

Phonos project⁷ 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

GiantSteps⁸ 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 is to create the "seven-league boots" for music production in the next decade and beyond,

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

⁷<http://phonos.upf.edu/>

⁸<http://www.giantsteps-project.eu/>

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 songs⁹ 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

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

⁹ "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.

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 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.

1.4.1 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* deal 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 artist, year of release, title, 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 context*. 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:

- 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*¹) 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. 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, 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 more complex. For this kind of systems, collaborative filtering data has shown to lead to better results [11] In this research field, the most used sources are mainly social tags and keywords extracted from webpages. The computation of similarity may happen through a Vector Space Model (a technique adapted from the Text-IR) or through co-occurrence analysis. In the next subsections we will see the characteristics and the performance of these two techniques.

¹<http://last.fm>

2.1.1 Computing Similarity With a Vector Space Model

2.1.2 Computing Similarity With Co-Occurrence Analysis

2.2 Audio Content Analysis

2.3 Conceptual Differences Between the Two Approaches

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

Chapter 3

Computing Music Similarity with Audio Content Descriptors

3.1 Literature Review

Part II

Methodology

Chapter 4

Phonos Catalogue of Songs

91 hours 43 mins 35 secs ꝑ length of entire catalogue

Chapter 5

Computation of Audio Features

5.1 Tools used for feature extraction, features extracted

5.2 Similarity computation (fast map)

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 Features

Write your Appendix content here.

Appendix B

List of Echonest Features

Write your Appendix content here.

Appendix C

Phonos: list of songs

Write your Appendix content here.

Bibliography

- [1] Avery 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.