Montevideo, 25 de setiembre de 2018

Señores Miembros de la Subcomisión Académica de Posgrado
Instituto de Ingeniería Eléctrica
Facultad de Ingeniería, Universidad de la República

PRESENTE

Por medio de esta nota avalo la presentación de la versión definitiva de la Tesis de Doctorado de Martín Rocamora. El documento incorpora las sugerencias y correcciones del Tribunal de Tesis.

Sin otro particular los saluda atentamente,

Luiz W. P. Biscainho
Universidade Federal do Rio de Janeiro

ISSN 1688-2784

Universidad de la República
Facultad de Ingeniería

COMPUTATIONAL METHODS FOR PERCUSSION MUSIC ANALYSIS:

the Afro-Uruguayan Candombe drumming as a case study

Tesis presentada a la Facultad de Ingeniería de la Universidad de la República por

Martín Rocamora

en cumplimiento parcial de los requerimientos para la obtención del título de Doctor en Ingeniería Eléctrica.

Directores de Tesis

Luiz W. P. Biscainho ...... Universidade Federal do Rio de Janeiro
Álvaro Pardo .... Universidad Católica del Uruguay
Luis Jure .... Universidad de la República

Tribunal

Andre Holzapfel (reviewer) ... Royal Institute of Technology (KTH)
Luiz Naveda (reviewer) ... Universidade do Estado de Minas Gerais
Gregory Randall .... Universidad de la República
Marcelo Fiori .... Universidad de la República

Director Académico

Pablo Musé .... Universidad de la República

Montevideo
April 6, 2018

COMPUTATIONAL METHODS FOR PERCUSSION MUSIC ANALYSIS: the Afro-Uruguayan Candombe drumming as a case study, Martín Rocamora.

ISSN 1688-2784

This manuscript was prepared in LaTeX using the class iietesis (v1.1).
It has a total of 207 pages.
Compiled on Tuesday 11th September, 2018.
http://iie.fing.edu.uy/~rocamora/phdthesis/

For all that music gives us.
For the spiritual and the mundane.
Its ability to bring and to
transcend the here and now.
The connection with life and beyond.

Acknowledgments

With a little help from my friends . . .

To the Afro-Uruguayans, for their gift and strength.

In memoriam Coriún Aharonián (1940–2017).

Abstract

Most of the research conducted on information technologies applied to music has been largely limited to a few mainstream styles of the so-called 'Western' music. The resulting tools often do not generalize properly or cannot be easily extended to other music traditions. So, culture–specific approaches have been recently proposed as a way to build richer and more general computational models for music.

This thesis work aims at contributing to the computer–aided study of rhythm, with the focus on percussion music and in the search of appropriate solutions from a culture–specific perspective by considering the Afro-Uruguayan *candombe* drumming as a case study. This is mainly motivated by its challenging rhythmic characteristics, troublesome for most of the existing analysis methods. In this way, it attempts to push ahead the boundaries of current music technologies.

The thesis offers an overview of the historical, social and cultural context in which *candombe* drumming is embedded, along with a description of the rhythm.

One of the specific contributions of the thesis is the creation of annotated datasets of *candombe* drumming suitable for computational rhythm analysis. Performances were purposely recorded, and received annotations of metrical information, location of onsets, and sections. A dataset of annotated recordings for beat and downbeat tracking was publicly released, and an audio-visual dataset of performances was obtained, which serves both documentary and research purposes.

Part of the dissertation focused on the discovery and analysis of rhythmic patterns from audio recordings. A representation in the form of a map of rhythmic patterns based on spectral features was devised. The type of analyses that can be conducted with the proposed methods is illustrated with some experiments.

The dissertation also systematically approached (to the best of our knowledge, for the first time) the study and characterization of the micro–rhythmical properties of *candombe* drumming. The findings suggest that micro–timing is a structural component of the rhythm, producing a sort of characteristic "swing".

The rest of the dissertation was devoted to the automatic inference and tracking of the metric structure from audio recordings. A supervised Bayesian scheme for rhythmic pattern tracking was proposed, of which a software implementation was publicly released. The results give additional evidence of the generalizability of the Bayesian approach to complex rhythms from different music traditions.

Finally, the downbeat detection task was formulated as a data compression problem. This resulted in a novel method that proved to be effective for a large part of the dataset and opens up some interesting threads for future research.

Resumen

La mayoría de la investigación realizada en tecnologías de la información aplicadas a la música se ha limitado en gran medida a algunos estilos particulares de la así llamada música 'occidental'. Las herramientas resultantes a menudo no generalizan adecuadamente o no se pueden extender fácilmente a otras tradiciones musicales. Por lo tanto, recientemente se han propuesto enfoques culturalmente específicos como forma de construir modelos computacionales más ricos y más generales.

Esta tesis tiene como objetivo contribuir al estudio del ritmo asistido por computadora, desde una perspectiva cultural específica, considerando el *candombe* Afro-Uruguayo como caso de estudio. Esto está motivado principalmente por sus características rítmicas, problemáticas para la mayoría de los métodos de análisis existentes. Así, intenta superar los límites actuales de estas tecnologías.

La tesis ofrece una visión general del contexto histórico, social y cultural en el que el *candombe* está integrado, junto con una descripción de su ritmo.

Una de las contribuciones específicas de la tesis es la creación de conjuntos de datos adecuados para el análisis computacional del ritmo. Se llevaron adelante sesiones de grabación y se generaron anotaciones de información métrica, ubicación de eventos y secciones. Se disponibilizó públicamente un conjunto de grabaciones anotadas para el seguimiento de pulso e inicio de compás, y se generó un registro audiovisual que sirve tanto para fines documentales como de investigación.

Parte de la tesis se centró en descubrir y analizar patrones rítmicos a partir de grabaciones de audio. Se diseñó una representación en forma de mapa de patrones rítmicos basada en características espectrales. El tipo de análisis que se puede realizar con los métodos propuestos se ilustra con algunos experimentos.

La tesis también abordó de forma sistemática (y por primera vez) el estudio y la caracterización de las propiedades micro rítmicas del *candombe*. Los resultados sugieren que las micro desviaciones temporales son un componente estructural del ritmo, dando lugar a una especie de "swing" característico.

El resto de la tesis se dedicó a la inferencia automática de la estructura métrica a partir de grabaciones de audio. Se propuso un esquema Bayesiano supervisado para el seguimiento de patrones rítmicos, del cual se disponibilizó públicamente una implementación de software. Los resultados dan evidencia adicional de la capacidad de generalización del enfoque Bayesiano a ritmos complejos.

Por último, la detección de inicio de compás se formuló como un problema de compresión de datos. Esto resultó en un método novedoso que demostró ser efectivo para una buena parte de los datos y abre varias lineas de investigación.

Contents

Chapter 1

Introduction

The work presented in this dissertation deals with music—more specifically, percussion music and musical rhythm—and is somehow at the crossroads of music-related technology, automatic analysis of music and computational musicology, while most tools, methods and concepts are drawn from signal processing, machine learning, statistics and information theory. This is probably to be expected, since modern research on music often transcends the boundaries of the scientific disciplines involved. The term "transdisciplinary" is sometimes used to describe the kind of approach needed—perhaps even more appropriate than the term "interdisciplinary"—which suggests that music is such a complex and multifaceted phenomena, that cannot be fully understood by a single discipline, or by different disciplines that are just put next to each other without much interaction [179].

Music is a fundamental aspect of human life, so ubiquitous and of such a paramount significance, that it can be even regarded as a human obsession [184]. There is no known culture, at present or in the past, that lacked a form of music, and some of the oldest physical artefacts ever build by archaic humans are musical instruments. All across the different cultures, it is a fundamental component of social activities; whenever people gather together there is music. The emotional power of music can touch us in profound ways, and constitutes a key factor in personal development [179]. Not surprisingly, music drives a strong economic sector within the cultural industries, which have gained importance in the 21st century and are becoming one of the most dynamic segments of the global economy [207]. In education, music is often a motivation for young people to develop interest in science and technology. Moreover, music is an excellent tool to promote respect for the diversity of social and cultural identity, and the care of cultural heritage [179].

It is therefore justified to say that music is more than just a domain of application [179]. To pose questions about such a fundamental human ability is to implicitly ask questions about perception, cognition, evolution, human interaction, culture and society. This means that understanding what music is, why we like it, and how we bond with it direct us to essential aspects of human nature [184].

The attribute of rhythm is of central importance in music. It concerns the way in which the musical events are arranged in time, grouped and organized, forming different structures and patterns. The topic has received increased attention in the past few decades, from scientific fields such as music theory, experimental psychology and cognitive science [193], resulting in novel theories of rhythmic organization [181]. This ultimately led to the implementation of computational models that are applicable in musicology, music search and discovery, and also to the study of music perception and cognition [280]. Nonetheless, so far, most of the research conducted has been largely limited to a few music traditions and repertoires, mostly from European and North-American art and popular music [36]. Since most of the approaches for computational rhythm analysis have not been developed taking into account a multicultural context, they do not generalize properly or cannot be easily extended to other music traditions [275].

This thesis work aims at contributing to the automatic and computer–aided study of rhythm, approaching some current research challenges mainly from a music technology background. It puts the focus on percussion music and in the search of appropriate solutions from a culture specific perspective by considering the Afro-Uruguayan *candombe* drumming as a case study. In this way, it also attempts to contribute to the overall field by identifying limitations of current methodologies and eventually providing some culture–aware insights that push ahead its boundaries towards the development of technologies that can deal with a wider range of music traditions. It also strives for addressing the issue of how these technologies could work as an effective tool leading to practical results that otherwise cannot be obtained. In turn, this could allow for cross cultural comparative studies and may help to preserve the diversity of our world's music. In the following, this chapter describes the research context and the motivation of the thesis. It also states the main contributions of the work and gives an outline of the text.

1.1 Research context

During the last twenty years, a new multidisciplinary field of research, known as Music Information Retrieval/Research (MIR), has emerged and steadily grown at the intersection of audio music processing, machine learning, music theory and musicology [262]. Primarily fuelled by the revolution brought by digital technology applied to music distribution and storage, it focuses on the processing of digital data related to music (such as editorial metadata, symbolic representations, scores, lyrics, and audio), and the development of methodologies to process and understand that data [266]. Most of the proposed methods and systems rely on audio content, which is extracted by means of signal processing techniques [213]. The gap between low-level descriptors extracted from audio signals (e.g. energy, spectral content), mid-level representations (e.g. note onsets, pitch estimates) and high-level musical concepts (e.g. meter, tempo, key), known as the "semantic gap" [68,179], is usually tackled by the application of machine learning.

The vast majority of the technologies and models developed in the field of MIR have been oriented towards mainstream popular music in the so-called 'Western' tradition,[1] which has certainly conditioned the problems addressed and the solu-

[1] In [275], the term Eurogenetic music is proposed as a way to avoid the misleading

tions obtained thus far. Although they proved to be effective for various music styles and repertoires, it seems that culture–specific approaches are needed to deal with other music traditions, such as those from Africa, China, India, or the Arab world. Fortunately, over the last few years there have been increasing efforts devoted to the study of traditional, folk or ethnic music [267]. More importantly, it has been gradually recognised in the MIR community that making sense of music is much more than processing audio files into high–level musical objects, and that the social, cultural and historical context in which music is embedded must be taken into account, as it highly influences how music is produced and perceived [267].

Therefore, a multicultural perspective is now being promoted in MIR research, in which other knowledge and methodologies, such as those from ethnomusicology, are coming into play [266]. Ethnomusicology is the study of music in its cultural context, considering what music means to its practitioners and audiences, and how those meanings are conveyed. It is highly interdisciplinary, encompassing several fields in the humanities and social sciences, such as cultural anthropology, folklore, performance studies, dance, cultural studies, gender studies, and race or ethnic studies. There are actually several terms to denote different and related approaches within musicology, like systematic musicology and empirical musicology [179]. Among them, computational ethnomusicology [286],[2] can be considered as the use of computational tools to assist in ethnomusicological research [291]. From this perspective, is ethnomusicology that could benefit from the advances provided by research on MIR and sound and music computing. Nevertheless, it can also be viewed as the connection of ethnomusicological concepts and frameworks with computational modelling, so as to enrich the current data–driven approaches of MIR with knowledge–based alternatives [291]. What is more, as noted in [137, 291], computer models can be regarded as 'theories' or 'hypothesis' about the problems tackled by ethnomusicologists, whose validity can be assessed empirically with large music collections at a scale not feasible by traditional means.

In recent years, the idea of "national rhythms", proposed by John Chasteen [71], has been given increasing importance in Latin American culture studies. It refers to forms of popular dance and music that emerged from the mixture of African and European dance and music traditions and that have gradually been assumed as symbol and expression of national identity [17]. The historical process by which they were created involved complex negotiations on issues of race, ethnicity, gender and social class, which represent the immense inequalities in status, well-being and power that have determined the processes of racial mixing and cultural creation throughout Latin America [16, 58]. Among those national rhythms, *candombe* drumming is one of the most characteristic elements of Uruguayan popular culture. Although it has been widely adopted by the society at large, it is deeply rooted in the Afro–Atlantic tradition and remains a symbol of the identity of the communities of African descent in Montevideo [110]. Internationally less known than other Latin American musics of African origin (such as Cuban

dichotomy of Western an non-Western music.

[2]Actually, according to [137], the term *computational ethnomusicology* can be traced back more than thirty years.

rumba and son, or Brazilian samba), Uruguayan *candombe* possess a considerable rhythmic wealth, which has influenced and was incorporated into various genres of popular music. In consequence, it deserves a thorough study to promote its wider recognition and the care of its cultural heritage.

1.2 Motivation

The importance of rhythm as fundamental dimension of music makes it a highly interesting topic of study. It is primarily an event-based phenomenon, so detecting and characterizing those events and discovering how they are organized in time into repeating structures and patterns, are important tasks that can be tackled using computational tools. Analysing and understanding these rhythmic patterns can provide useful insights into the music being studied and may help to extract semantically meaningful higher level concepts. For instance, the rhythmic patterns can be indicative of the underlying musical structure of a musical piece.

Furthermore, the tools developed for rhythm analysis can be further used in several practical applications. These include digital audio workstations for music editing and processing, DJ-mixing software and hardware products, and other tools for creating music or performing live. With the need to handle large collections of digital music, automatic tools for rhythm analysis can also be applied to problems such as content based music retrieval, enhanced navigation of music collections and intelligent music archival and organization. Consequently, the target users of these technologies encompass from casual music listeners to musicians, music students and teachers, including also musicologists and cognitive scientists, sound engineers and music producers, as well as record labels and cultural institutions.

The study of percussion music provides a suitable scenario for bringing into focus the rhythmic aspects of music, without the need of considering the interplay with other music dimensions, such as melody and harmony. The selection of *candombe* as a case study is motivated—apart from a musicological interest, and because of its cultural relevance—by the fact that some of its characteristics are troublesome for most of the existing rhythm analysis methods. The identification of challenging music styles and the development of style-specific algorithms for rhythm analysis is a promising direction of research to overcome these limitations.

In addition, the study of a particular music tradition outside the Western-centred paradigm can help to build richer and more general models than the ones that currently dominate the research on information technologies (IT) applied to music [267]. At present, there is a significant gap between the current capabilities of the music technologies used in commercial services and the needs of our culturally diverse world. For example, during the last few years we have witnessed the upsurge of music recommendation services which make extensive use of automatic tools (e.g. Last.fm, Pandora, Spotify). But if the efforts aimed at the development of IT applied to music are exclusively market–driven and do not take into account a multicultural reality, the progress made can actually deepen the existing bias towards the dissemination, recommendation and access to a very reduced type of music, restricting the diversity of the offer [267].

In addition to the above, to the best of our knowledge, the research work described in this dissertation is the first one to undertake the analysis of *candombe* drumming using computational tools. For this reason, it is impelled by the goal of producing relevant results and useful resources, so as to open up the way for further research on the topic. At the same time, the opportunity to help bring the attention of the research community to this rich music tradition, and the chance to work in close collaboration with highly accomplished musicians and passionate scholars, is a very rewarding task.

1.3 Goals and scope of the thesis

The goal of this thesis is to build musically informed and domain specific signal processing and machine learning methods for rhythm analysis, oriented towards Afro-Uruguayan *candombe* drumming. For this reason, it aims to bring in as much musical knowledge to the methods as possible, so that the extracted information is musically relevant and useful. In order to accomplish that, the work presented relies on the musicological literature available and borrows from consultation with music experts and renowned *candombe* drummers. In this respect, the research work was supervised by Luis Jure, who has been involved in the study of *candombe* drumming from a musicological perspective since the early 1990s.

The dissertation includes examples of the type of analyses that can be performed with the proposed computational methods, which hopefully may help to obtain a better understanding and provide deeper insights into the nature of *candombe* rhythm. However, the thesis does not aim to make any significant musicological conclusions. In addition, the analysis methods developed in the thesis in no way aim to replace expert musician opinions.

An audio recording is the primary source of information considered and the algorithms developed are intended to work on real world representative music collections. Despite the fact that some music generation tools are implemented, and that synthetic rhythmic patterns and performances are rendered as audio files for experimental purposes, the thesis focuses only on music analysis and not on music generation, composition and synthesis.

The broad goals of the research can be summarized as follows:

1. To build annotated music collections of *candombe* drumming (both audio and video), useful for future research.

2. To identify challenges and opportunities in automatic rhythm analysis of *candombe* drumming and formulate relevant analysis problems.

3. To convert those problems into engineering formulations amenable to quantitative and qualitative analysis using signal processing and machine learning.

4. To propose and implement computational tools for rhythmic pattern discovery and analysis for *candombe* drumming.

5. To develop novel methods for rhythm analysis capable of tracking beats and downbeats in *candombe* drumming.

Another goal of the thesis is to develop open and reproducible research. Hence, all the data and code used in this work will be made available to the research community through open source platforms under open licenses. Whenever possible, resources will be provided to reproduce the presented results.

1.4 Publications

The following is a list of publications and other forms of relevant scientific communication that were produced in the context of this thesis work.

1.4.1 Complete papers in peer–reviewed conferences

[254] Martín Rocamora, Luis Jure, and Luiz W. P. Biscainho. Tools for detection and classification of piano drum patterns from candombe recordings. In *Proceedings of the 9th Conference on Interdisciplinary Musicology (CIM 2014)*, pages 382–387, Berlin, Germany, 4-6 dec. 2014. iie.fing.edu.uy

[219] Leonardo Nunes, Martín Rocamora, Luis Jure, and Luiz W. P. Biscainho. Beat and downbeat tracking based on rhythmic patterns applied to the Uruguayan candombe drumming. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015). Málaga, Spain*, pages 264–270, 26-30 oct. 2015. ismir2015.uma.es

[252] Martín Rocamora and Luiz W. P. Biscainho. Modeling onset spectral features for discrimination of drum sounds. In *Proceedings of the 20th Iberoamerican Congress on Pattern Recognition (CIARP 2015). Montevideo, Uruguay*, pages 100–107, 9-12 nov. 2015. link.springer.com

[197] Bernardo Marenco, Magdalena Fuentes, Florencia Lanzaro, Martín Rocamora, and Alvaro Gómez. A multimodal approach for percussion music transcription from audio and video. In *Proceedings of the 20th Iberoamerican Congress on Pattern Recognition (CIARP 2015). Montevideo, Uruguay*, 9-12 nov. 2015. link.springer.com

[255] Martín Rocamora, Luis Jure, Bernardo Marenco, Magdalena Fuentes, Florencia Lanzaro, and Alvaro Gómez. An audio-visual database of candombe performances for computational musicological studies. In *Proceedings of the II Congreso Internacional de Ciencia y Tecnología Musical (CICTeM 2015)*, pages 17–24, Buenos Aires, Argentina, 26-28 sep. 2015. iie.fing.edu.uy

1.4.2 Extended abstracts in peer–reviewed conferences

[159] Luis Jure and Martín Rocamora. Microtiming in the rhythmic structure of candombe drumming patterns. In *Fourth International Conference on Analytical Approaches to World Music*, New York, USA, June 8-11 2016.

[160] Luis Jure and Martín Rocamora. Clave patterns in Uruguayan candombe drumming. In *16th Rhythm Production and Perception Workshop*, Birmingham, UK, July 3-5 2017.

1.4.3 Other presentations in workshops and conferences

[253] Martín Rocamora and Luis Jure. Rhythmic pattern tracking: the Uruguayan candombe drumming as a case study, September 26–28 2013. Oral presentation at the *I Congreso Internacional de Ciencia y Tecnología Musical (CICTeM 2013)*, IUNA. Buenos Aires, Argentina.

[250] Martín Rocamora. Automatic analysis of percussion music: the Afro-Uruguayan candombe drumming as a case study, October 12–15 2014. Short oral presentation at the *II International Workshop on Cross-disciplinary and Multicultural Perspectives on Musical Rhythm and Improvisation*, New York University Abu Dhabi. Abu Dhabi, United Arab Emirates.

[251] Martín Rocamora. Interpersonal music entrainment in Afro-Uruguayan candombe drumming, July 13–19 2017. Oral presentation at the *44th International Council for Traditional Music (ICTM) World Conference*. Limerick, Irleand. Work in collaboration with Nori Jacobi, Rainer Polak and Luis Jure.

1.5 Organization and outline of the thesis

The rest of the manuscript is organized as follows.

Chapter 2. Afro-Uruguayan Candombe The aim of this chapter is to give an account of the main historical and cultural aspects of the practice of *candombe*, including a description of the most relevant musical traits of its rhythm and the way it is performed.

Chapter 3. Data collection and generation The chapter is devoted to the description of the data and music collections used in this thesis work. It also presents a dataset of labelled *candombe* recordings for beat and downbeat tracking and an audio-visual database of *candombe* performances, both created during the development of the research described in this dissertation.

Chapter 4. Audio features In this chapter, the audio features used throughout the thesis are presented. The audio feature extraction process and some representations built upon it are described. The usefulness of the features is assessed in the context of onset detection and classification of *candombe* drum sound events.

Chapter 5. Analysis of rhythmic patterns In this chapter some techniques are proposed for the detailed analysis of the rhythmic patterns of *candombe* drumming from audio recordings. A set of tools is proposed for the study of rhythmic patterns that span over the four-beat cycle, with the aim of investigating its different types and forms. Additionally, some experiments are proposed in order to assess the exact nature of the micro-temporal deviations found in the rhythmic patterns.

Chapter 6. Beat and downbeat tracking The chapter deals with the development of a beat and downbeat tracking algorithm suitable for *candombe* drumming recordings. A supervised scheme for rhythmic pattern tracking is proposed, which aims at finding the metric structure from an audio signal, including the phase of beats and downbeats.

Chapter 7. Analysis based on information theory The main motivation of this chapter is to recast the down-beat detection task as a data compression problem. For this purpose, a lossy compression framework based on the rate-distortion theory is adopted. Additionally, it turns out that the obtained description is well suited for addressing other related tasks, namely the assessment of performances' complexity and the estimation of the number of different rhythmic patterns in a given recording.

Chapter 8. Conclusions The document ends with a critical discussion of the thesis work, including some directions for future research.

Chapter 2

Afro-Uruguayan Candombe

2.1 Introduction

Arguably the most widespread vision of Uruguay—both from inside and abroad—is that of a country of white population comprised of descendants of European immigrants, mainly from Spain and Italy [23, 64]. This was a praised attribute of national identity for more than a century—since the late 1800s and early 1900s—that supposedly differentiated the country favourably from the rest of Latin America, at least in the eyes of the dominant elites [18, 64]. This homogeneous self-portrait of the Uruguayan population is said to have contributed to social cohesion, but more than likely at the expense of neglecting and excluding some of the cultural legacy of minorities [64]. It is only in the last few decades that the diversity of the Uruguayan population is being widely recognized, and the role of various ethnic groups is being thoroughly studied [24, 64, 116, 211, 241, 242].

During the period of Spanish colonial rule, thousands of enslaved Africans were brought to the land corresponding to Uruguay, formerly known as *Banda Oriental*—the eastern shore of the Uruguay river. Africans and their descendants took active part in the process of independence and played important roles in the national life of the country. Being an heterogeneous community, they gathered in numerous social and civic organizations, and following the abolition of slavery they created the second-largest black press in Latin America and a racially defined political party [18]. Afro-Uruguayans were also key contributors to the shaping of the popular culture of the country and two of its most important musical forms, *tango* [127] and—to a greater extent—*candombe* [106].

Nowadays, the practice of *candombe* is one of the most characteristic and defining features of Uruguayan popular culture, and while still being primarily associated with the Afro-Uruguayan community, it has long been adopted by the society at large. At the same time, it is probably the most obvious local sign of African ancestry, linking Afro-Uruguayans to the so-called Afro-Atlantic diaspora [123]. The fundamental component of this tradition is the *candombe* drumming, performed by groups of drums playing a distinctive rhythm. All along the year, specially on weekends and public holidays, players meet at specific points to play *candombe* marching on the street, as shown in Fig. 2.1. Since 1956, the municipal government of Montevideo—Uruguay's capital city—organizes during Carnival a major event convoking thousands of people, named the *Desfile Oficial de Llamadas*, a parade of *candombe* groups called *comparsas*, comprising not only drums but also dancers and traditional characters in costume. In the last decade of the 20th century, *candombe* drumming grew in popularity, and today groups of performers can be found all over Montevideo, as well as in other cities of the country, which also celebrate highly attended events during Carnival. Its rhythm was also integrated in different ways into several genres of popular music, like tango rioplatense, *canto popular* (folkloristic popular song), beat/pop/rock in the so-called *candombe beat*, etc. In 2009, *candombe* was inscribed on the Representative List of the Intangible Cultural Heritage of Humanity by UNESCO, taking into account, among other things, that it is "a source of pride and a symbol of the identity of communities of African descent in Montevideo" [287].

Figure 2.1: Group of Candombe drummers. (*Mario Marotta, reproduced with permission*)

In order to acknowledge this cultural context and its origins, the term Afro-Uruguayan is adopted in this work to refer to *candombe*. This is common practice in the musicological realm in Uruguay since the pioneering work by Lauro Ayestarán (1913-1966) [27–31], who designated some music and instruments as being Afro-Uruguayan, and also used the denomination Afro-Montevidean to circumscribe them only to the capital city. Besides, this kind of terms, such as Afro-descendent, are promoted by a large part of the black movement not only in Uruguay but throughout Latin America [18, 247, 256]. It is an attempt to replace the language of race, based on phenotypes—e.g. colour or facial features—with the language of ethnicity—based on e.g. common ancestry, religion, place of origin or other cultural heritage [18]. Therefore, it is applied to those individuals considered by themselves or by others to be of 'black' African ancestry, as well as to their cultural elements and practices [16]. The motivation of this preference for ethnic over race terminology is based on the present understanding of race strictly as a social construct[1] that segments the human population into ranked categories that defines people as superior or inferior according to their phenotypes [243].

In short, *candombe* drumming and its related practices play key roles in Uruguayan popular culture. On the one hand, *candombe* has been widely adopted as a *national rhythm*[2] [18, 71], and is regarded as an identity mark by a large part of the population. On the other hand, since it was mainly developed by an ethnic-racial group under domination and later subordination, it grew into a sign of cultural resistance which gained paramount importance for African descendants, even nowadays. Somewhat paradoxically, this profound cultural meaning is often overlooked by both the government and large part of the society, binding *candombe* only to Carnival and show-business, which ultimately contributes to social impairment of Afro-Uruguayans and to racism [103, 109, 247, 256].

The study of *candombe* from a musicological perspective dates back to the pioneering work by Lauro Ayestarán in the 1950s [27–31], followed by his disciple Coriún Aharonián [5–7]. Several years later, a new generation of researchers focus on *candombe* by the late 1980s and early 1990s [104, 105, 124, 156, 157]. Following this studies, the book by Luis Ferreira is considered a milestone [106]. Subsequent musicological work produced by the same authors include [107, 108, 110–112, 125, 126, 158, 235–237]. Nevertheless, considering a broader scope, the early scholarly literature on Afro-Uruguayan history and culture also involves [84,85,148,206,234, 246, 259, 288]. Fortunately, the amount of Afro-Uruguayan studies has undergone a steady increase in the last two decades, see for instance [13,18,22,55–60,69,116, 127,128,185,211,212,222,223,227–229,242,256,260]. This is partly in response to debates and discussions promoted by Afro-Uruguayan activists and intellectuals [109, 211, 212, 256], and partly as a result of the increasing interest in Afro-Latin

[1] The idea of race, in its modern meaning in the Occident, was mainly developed by Western Europeans following their global expansion, as a way of granting legitimacy to slavery and social domination. To most contemporary scientists, the concept of race as a biological human category is baseless [243]. However, it has proven to be one of the most effective and long-lasting instruments of universal social domination, giving rise to racism [115], and with profound effects on social relations and cultural features up to present times [265]. The relationship between race and ethnicity can be misleading, and often both categories are used interchangeably, though probably with different connotations. In fact, three different kinds of social groups can be identify as potential victims of discrimination an prejudice based on race [264]. Firstly, those individuals that combine a phenotype sign as a racial marker with a distinctive cultural heritage, which constitutes an ethnic-racial social group. This is, for instance, the case of Afro-Uruguayans involved in *candombe* practices. Then, there can be also race without ethnicity, such as in people with a phenotype associated to race but without carrying a differentiated cultural legacy. Finally, ethnicity without race, that is people without distinctive phenotypes, but still practitioners of some idiosyncratic cultural tradition.

[2]The idea of *national rhythm*, proposed by John Chasteen [71], has gained increasing importance in Latin American cultural studies in recent years. It denotes forms of popular dance and music, arising from the mixtures of African and European traditions. Such rhythms (Argentinean and Uruguayan tango, Brazilian samba, Colombian cumbia, Cuban rumba and son, Dominican merengue and bachata, and Puerto Rican bomb, plena and salsa) have been gradually assumed as a symbol of national identity [17].

American studies worldwide [16, 18, 185]. Not surprisingly, the vast majority of all this literature has been written and published in Spanish.[3]

The aim of this chapter is to give an account of the main historical and cultural aspects of the practice of *candombe*, including a description of the most relevant musical traits of its rhythm and the way it is performed. Next section offers a historical perspective that hopefully contributes to understanding contemporary *candombe* tradition in the light of its earliest phases and subsequent evolution. Then, Section 2.3 deals with *candombe* drumming, describing the drums and its rhythmic and metric structure. It also gives a brief overview of the main influences of *candombe* drumming into popular music.

2.2 Historical perspective

A long historical process lead to model this cultural expression as it is known nowadays [18], since *candombe* drums have been played for almost two centuries in Montevideo. Firstly, from the end of 18th century up to the abolition of slavery in the second half of the 19th century, intercultural transformation processes took place among different African cultures (mainly West African Mahi and Nago, Congo-Angolan and Mozambican) [110]. In the remainder of the 19th century the African-based cultural practices were re-elaborated. This was mainly determined by the modernization processes imposed by the white elites since the constitution of Uruguay as a nation, and by the steady raise of the number of European immigrants, primarily from Italy and Spain. During the 20th century, Carnival associations were created within neighbourhoods of African descent families (also integrating European immigrants), the parades were turned into core symbols of cultural identity promoted by the national and municipal authorities, and some artists linked to *candombe* reached massive success.

The socio-economic characteristics of the process of colonization and later development of the country made that the contribution of Afro-descendants to the national culture exhibits more visibility and definition in the urban environment [60, 116]. Although this region did not give rise to a plantation society as in most areas of the Americas, during the late 18th century, a large black urban community and the typical social life of other slave trading ports developed in Montevideo[4] [58]. Today, the proportion of the population that acknowledges

[3]Among the few exceptions accessible to English-language readers are: the chapter by Tomás Olivera Chirimini [222], which gives a brief historic overview of the cultural practices of communities of Black African descent in Uruguay, the book by George Reid Andrews [18], which offers a comprehensive history of Afro-Uruguayans from the colonial period to the present, the book by Alex Borucki [58], which studies the lives of Africans and their descendants in Montevideo (and Buenos Aires) from the late colonial period to the first decades of independence, and the paper by Luis Ferreira [110], which deals with *candombe* drumming, focusing on the body movements that generate sound and music.

[4]Slaves were the major part of the labour force of the rural and urban economy of the territory. Their work was used in almost all the activities, from the colonial time, during the revolutionary period and after the establishment of the nation state. They were

African ancestry in Uruguay is about 9%, being the largest ethnic minority [65]. Although the highest proportion of Afro-descendants is found in the northern border departments of Rivera and Artigas (about 17%, probably due to the influence of Brazil, the northern neighbour country) [65], these zones are sparsely populated and most Afro-Uruguayans live either in Montevideo or the adjacent suburban department of Canelones [18, 65]. Besides, *candombe* has been mainly associated with the capital city, despite its dissemination to other cities in recent years [237]. By 1965, Lauro Ayestarán restricted its existence only to Montevideo, acknowledging some lost traditions in other cities of the country in the past [30,31]. Thus, Montevideo is the main focus of this chapter.

2.2.1 Africans in Montevideo: slave trade and African origins

The early records of the presence of enslaved Africans in the Banda Oriental—the territory of what is today Uruguay—date back to the beginning of the 17th century [60,222]. Its first city, Colonia del Sacramento, was founded by Portuguese military forces in 1680. From its port, goods and slaves were smuggled to the other shore of the Río de la Plata, into Buenos Aires—the principal city of the region, established in 1580 under the Spanish colonial rule [57]. The city of Montevideo was founded between 1724 and 1730, as part of Spain's effort to prevent Portuguese incursions into the Río de la Plata [18, 222]. Some years later, the authorities of the city ask permission for the introduction of enslaved Africans to serve as labour force, and since 1743 the trade of slaves to Montevideo began to run on a regular basis [60]. From 1776 to 1814, Spain organized this part of its empire in the Viceroyalty of the Río de la Plata, which roughly included the modern territories of Argentina, Bolivia, Paraguay and Uruguay—with Buenos Aires as the capital.

During the second half of the 18th century, Montevideo turned out to be the port of entry for ships sailing to and from Buenos Aires and the base of the Spanish navy in the South Atlantic [57]. In 1791, the Spanish Crown declared Montevideo the only authorized entry for slaves to the Río de la Plata, Chile and Perú [57]. The slave trade thus became one of the most important economic activities of the city, including the earnings due to commercial taxes [60]. Merchants of Montevideo developed busy networks with Luso-Brazilians and Portuguese, the most experienced slave traders of the South Atlantic [57]. The trade of cattle hides—the main product of the region—with Brazil was authorized, and upon their return the ships could bring slaves. In this way, Brazil became the main supplier of slaves to the port of Montevideo [60]. The slave trade to the Río de la Plata was then built upon complex trans-imperial networks of commerce and smuggling involving primarily the cities of Buenos Aires, Montevideo, Rio de Janeiro and Salvador [57].

At least 70,000 enslaved Africans arrived in the Río de la Plata between 1777 and 1812, following straight trade routes from Africa and also indirectly through Brazilian ports [57], see Fig. 2.2. This constitutes the most important demographic

employed in all sort of roles of Montevideo's economy, such as street vendors, laundresses, seamstresses, stevedores in the city port, domestic servants for well-off families, and also practised several professions, like shoemaking, carpentry, masonry and blacksmithing [116].

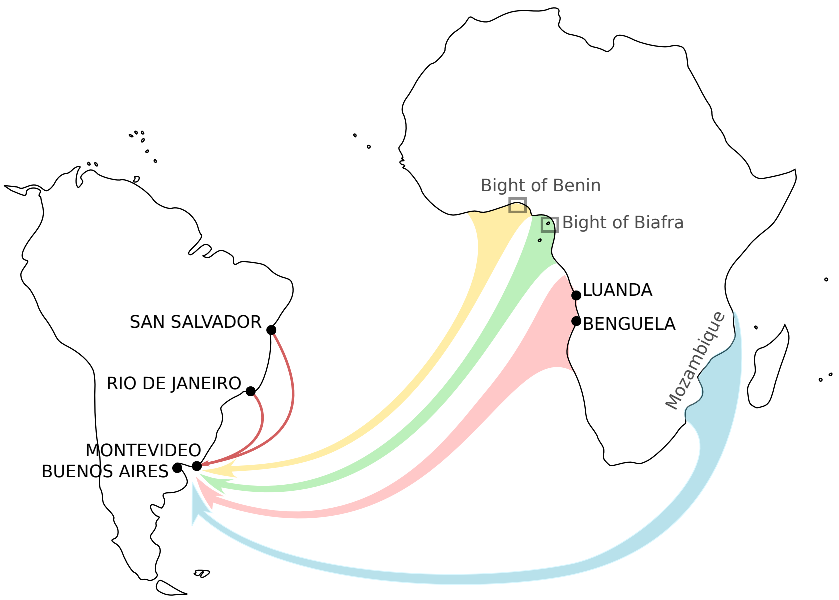


Figure 2.2: Slave trade routes from South America and Africa arriving at the Río de la Plata.

event since the Iberian colonization to this region, given that Buenos Aires had only 43,000 inhabitants by 1810 and Montevideo no more than 12,000 by 1803 [57]. During the first decades of the 19th century, about thirty percent of the population of Montevideo was enslaved [57, 116]. In spite of the several difficulties for determining the exact origin of the enslaved Africans in the Americas—for instance, the pervasive smuggling, errors and omissions of trade records, departure ports receiving people from different ethnic groups—there is a broad agreement on the large diversity of the people arriving to the Río de la Plata, coming from many different regions, and that predominantly belong to the vast Bantu cultural area [57, 106, 116, 222, 260].

While all broad areas supplying slaves to the Americas took part in the direct traffic to the Río de la Plata, three regions provided the large majority (probably about 85%), namely South-East Africa (Mozambique), West-Central Africa (Loango and Angola), and the Bight of Biafra (also known as the Bight of Bonny) [57]. The remainder portion of the direct trade to the Río de la Plata (about 15%) embarked from the Bight of Benin, Upper Guinea and Gold Coast [57].

But the slave trade from Brazil to the Río de la Plata was actually larger than the direct trade from Africa (60% between 1777 and 1812), with more than half of the people embarking in Rio de Janeiro and almost thirty percent in Salvador [57]. Nearly all slaves coming from Rio de Janeiro were most likely from Angola originally—embarking in West-Central Africa mainly from only two ports: Luanda and Benguela—while a minority, particularly those coming from Salvador, may have initially embarked in the Bight of Benin [57]. In this way, more people from West-Central Africa and the Bight of Benin arrived to the Río de la Plata coming from Rio de Janeiro than directly from Africa [57].

Several of these origins are reflected into the names of the African-based associations that existed during the late colonial period and throughout most of the 19th century in Montevideo (and Buenos Aires), which functioned as mutual aid societies and organized social activities that helped to preserve part of their traditions. Their names referred to ethnic groups (e.g. Mandinga, Mahi), points of embarkation in Africa (e.g. Mina, Benguela), and large geographical areas (e.g. Angola, Congo and Mozambique) [57, 222].

2.2.2 African-based associations: naciones and cofradías

The uprooting and forced cultural assimilation experienced by Africans and their descendants, prompted processes of syncretism, recreation and resistance, which responded to the need to establish themselves as a social group [60]. In cities all over the Americas, enslaved Africans and free blacks[5] formed different kinds of associations, that contributed to define their sense of belonging and identity. From the late 18th century until the end of 19th century, lay brotherhoods called *cofradías* and African-based fellowships called *salas de nación*, were the most important of these associations in Montevideo [58, 126]. Although they were forms of social control somehow—as they were strictly supervised, or even promoted, by colonial and later national authorities—, they allowed Africans and their descendants to meet and recreate part of their cultural practices.

As in the rest of the Spanish colonies, an evangelizing effort of the Catholic church—intimately associated with the colonial power—was to organize Africans into lay brotherhoods. In Montevideo, there were at least two important *cofradías* devoted to black saints; one worshipped Saint Benedict of Palermo, and the other Saint Balthazar.[6] Founded in 1773 and 1787, respectively, the former functioned until approximately 1892 [126]. It has been suggested that the actual meanings attributed by Africans in the Americas to the representations of the saints were related to their own spiritual traditions, instead of those intended by the Catholic church [222]. Thus, in the context of Christian rituals, such as saints' feast days or the Corpus Christi, they were able to express some of their own religious and festive behaviours in public processions with music and dance of African origin [222].

However, from a cultural production and reproduction perspective, the most important associations were the *salas de nación*: mutual aid societies based on African ethnic identities[7] [60]. The records and chronicles document the existence of African-based *nations* since the late 18th century in the Río de la Plata [58], coming to about twenty simultaneously functioning in Montevideo from early to mid 19th century [60, 126, 130, 212]. Apart from providing social networks and support to make life endurable, they fulfilled a religious function. The passage from life to death was of central importance, so people from all *nations* used to

[5]Sometimes slaves worked for other people receiving a pay. The money obtained was mostly for the master, except for that generated on Sundays and holidays, which could be retained by the slaves and saved in order to buy their freedom or that of their family. In other cases the manumission could be obtained from the master as a grace [116].

[6]A women's *cofradía* devoted to Our Lady of the Rosary is also documented [127].

[7]Resignified by both continuities and ruptures in social networks and experiences [58].

attend wakes and funerals of their members. Actually, African meetings in the Río de la Plata were firstly denominated *tambos*, after a funeral ritual from Portuguese Angola [58]. Dance and music played a prominent role in these rituals, and constituted the means to preserve some of their traditions, such as the use of native languages in songs [106]. They also held weekly meetings—on Sundays—for drumming and dancing [89], as well as on religious and public holidays [18]. Although the actual religious meaning of some of these rituals is missing from surviving sources [58], they were powerfully spiritual events, deeply rooted in African religious observances and beliefs [18, 106].

The different ceremonies took place either indoors, in closed spaces called *salas*, or outdoors, in open areas called *sitios* and along the streets in the form of processions [106]. Firstly, *salas* were located within the walled city toward the south, and *sitios* were vacant areas to the southern coast of the city in the Cubo del Sur and in the Market square in the centre of town [222]. But during the second half of the 19th century, African-based associations and celebrations were displaced by the authorities to the south of the new city and the Cordón district [58, 126]. Large part of the African and Afro-descendant population was settled there, and the southern coast developed what currently constitute the historic black neighbourhoods of Montevideo: Barrio Sur and Palermo [58], see Fig. 2.4.

In the late colonial Río de la Plata, the African-based celebrations, involving dancing, singing and drumming, were known generically as *tangos* and *tambos*—as previously noted [18, 28, 127]. The terms were used to designate both the meeting places and their music and dancing. But a new name for these meetings, namely *candombe*, became widespread in press and official records in Montevideo in the 1830s,[8] and it can be assumed that was in oral circulation since approximately a decade before [58]. In fact, the term *candombe* was previously in use in Rio de Janeiro (not to be confused with *candomblé*), appearing prominently in police records of early to mid 19th century, to designate meetings with African dance and music—and used interchangeably with the term *batuque*[9] [58]. Just before the appearance of the word in written records in Montevideo, the city was under Luso-Brazilian occupation (1817-1829), thus connected to Rio de Janeiro probably as never before [58]. This fact fuels the hypothesis that Rio de Janeiro could have functioned as the centre from which West Central African-based rituals were modified before turning to other regions, such as the Río de la Plata and Minas Gerais[10] [58].

[8]The term *candombe* first appeared in writing in 1829, in a judicial case concerning the killing of a soldier who had been watching a *"candombe de los negros"* during Carnival, in the south of Montevideo [58]. Later it appeared in a local newspaper, in 1834 denoting the dances held by the *nations* in Saturdays and holidays, and in 1835 in the verses of a song in the *bozal* dialect commemorating the 10th anniversary of the freedom of wombs law, written by white poet Francisco Acuña de Figueroa [18]. Police records also show use of the term, for instance, in two attempts to regulate the *nations'* meetings in 1839 [28, 126].

[9]The term *batuque* was also in use in Montevideo and refers to a wedding dance in Angola, also emphasizing the West Central African predominance of ritual origins [58].

[10]A tradition called *candombe* also developed in late colonial Minas Gerais and it is still practised today. It was a secret society within the Catholic black brotherhood of Our

Regarding the origins of the word *candombe*, some bibliography sources point to the Central Africa Kimbundu word *ndombe*, the ethnonym denominating the Ndombe people, the largest group that lived along the Atlantic coast near Benguela, when the city was founded by the Portuguese. Together with the qualifier *ka*, meaning 'pertaining to' or 'the location of' as a place, they formed the word *candombe*, which could probably indicate things having to do with Ndombe people [58, 212, 222].[11]

Due to the cultural diversity of the African *nations*, there were in fact a lot of different practices, for instance, different dances (e.g. calenda, bámbula and chica), songs, rhythms and musical instruments (e.g. mazacalla, marimba, tacuara, porongo and drums) [28, 126, 127, 222]. Apart from the Sunday meetings, the *nations* organized processions in which they visited each other on the occasion of mourning or feast day. The most important *candombes* took place during the *Candombe de Reyes* (Candombe of the Kings), which was a very common festivity all throughout the Americas [126]. It was the annual celebration of the coronation of the Kings and Queens of the Congo and Angola, beginning on Christmas Days and culminating on the Day of the Kings, January 6, which for some Afro-South American communities was the feast day of Saint Balthazar [222]. The *salas de nación* were internally organized in the manner of a monarchy, and the presiding king and queen were selected in this occasion from among the most respected elders of the community [222]. The festivity included a mass at Montevideo's cathedral, followed by a parade to the southern part of the city, which ended in an outdoor dance [126], see Fig. 2.3. The monarchs headed the procession, which was leaded by the master of ceremonies, namely the *bastonero*, followed by a cortège of men and women dressed as elegantly as possible, after whom came the musicians playing drums and other instruments. Members of the black militias were solemnly in their military uniforms, and some officers were also kings of the African *nations* [58]. Although the city council banned the dances within and outside the walls on various occasions, they were one of the most important public events of Montevideo, extensively mentioned in the local press. By the 1850s and 1860s, the Day of the Kings was attended by thousands of people, about 10 percent of the city population, including many white people [18, 58].

Conversely, the references to the rituals held indoors are scarce, and some *nations* often adopted the form of initiatic secret societies [106, 222]. State authorities were frequently uneasy about these gatherings, their repression and control obviously contributing to the obscurity and secrecy. Certainly, one of the main purposes of the *salas de nación* was to defend the rights and interest of their community. For instance, they helped to obtain the manumission of their kings or members that were to be sold by their masters [60]. They also provided some kind

Lady of the Rosary, whose members were devotees of the three drums of *candombe* [165]. However, it exhibits no other apparent connections to current practices in Montevideo [58].

[11] However, it has been also argued that the words *ka* and *ndombe* actually come from two different languages [58]. Another option is provided in [165], concerning *candombe* in Minas Gerais, where the Bantu origins of the word are drawn from linguistic studies, meaning 'to pray' or 'to ask the intercession of'.

of political representation and lobbying organizations [18]. In their processions on major religious and civil holidays, the monarchs of the *nations* would pay a visit to the authorities to convey their greetings, assuming positions of dignity and influence denied to them in daily life, while asserting their collective presence and their Africanity through music and dance [18]. Their acts of resistance were usually in peace, but also led to some revolts, such as the attempt to establish a maroon community in 1803 and the free blacks and slaves conspiracy in 1833 [58,129,222].

With the death of the last people brought from Africa—many of them in the struggles of the 19th century—and with the process of secularization and positivism promoted following the creation of the nation state [33], the *salas de nación* gradually disappeared towards the end of the 19th century—the last references to them in press occurring during the 1880s [126]. However, some of its practices may have remained in the private sphere and some of its elements survived in new cultural expressions, namely *sociedades*, *comparsas* and *conventillos* [18,106,127].

Figure 2.3: *Candombe*, painting by Pedro Figari (1861–1938), oil on cardboard, 1932. It belongs to a series in which the author portrayed the festivities as he witnessed as a child.

2.2.3 Uruguay as a nation: sociedades, comparsas, conventillos

In Uruguay, as happened in most of Spanish America, warfare and the foundation of the republic roughly coincide with the end of the slave trade[12] and the abolition of slavery [58]. One of the principal points of leverage was the need for soldiers to battle the independence and civil wars [18].[13] Yet, the abolition of slavery was actually a long process that extends from the first laws of the independence (1825) to the end of the *Guerra Grande* civil war (1839-1851) and beyond [116].[14]

[12]The last trans-atlantic slave trip direct from Angola arrived in Montevideo in 1835 [58].

[13]In the Río de la Plata, colonial black militias can be traced back to the end of the 18th century and were open only to free blacks. But when the revolution began in 1810 in Buenos Aires, the recruitment of slaves by emancipation became prevalent. Freed blacks joined forces on all sides of the armed conflicts, either willingly or being forced [58].

[14]In 1825 the freedom-of-wombs law was approved. In 1837 a law was passed on the prohibition of the slave trade. During the *Guerra Grande*, both sides of the conflict pro-

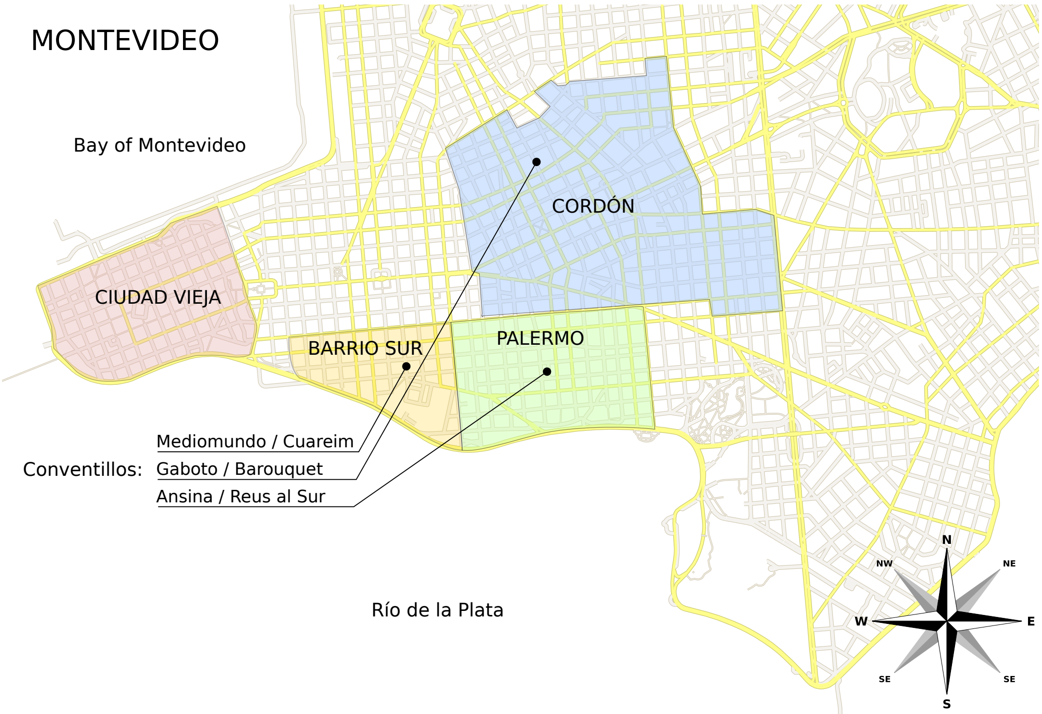


Figure 2.4: Montevideo city. The old walled city, the traditionally Afro-Uruguayan neighbourhoods, and the location of the three most significant *conventillos* are shown.

However, the ruling class was unwilling to lose slaves' workforce, so the laws were not fully applied and new ways of entering slaves were devised.[15] Slavery-like practices continued throughout the 19th century and still in the 20th century, like granting of employment in exchange for housing, food and, in some cases, education [56, 59].

Men and women of African origin or ancestry had to insert themselves as free people in society and to establish labour relations based on new rules. The large numbers of European immigrants that the nation received between 1880 and 1930 also shaped social relations and job opportunities decisively.[16] Thus, black people took on less paid and informal jobs, or were directly out of the job market [116, 127]. The processes of discipline imposition, which since the second half of the 19th century fell on the working classes, particularly affected the black population, from institutions such as the army or school, or the regulations of labour

claimed abolition, the *Colorados* in 1842 and the *Blancos* in 1846, followed by compulsory enrolment of all able-bodied man of African ancestry into the army. Children and women remained under the authority of their former masters in "custody" or as "apprentice". In 1853 the custody of the sons of the slaves emancipated by the abolitionist laws was ceased. By marrying, women were freed from all legal links with their former masters [56, 60].

[15] After abolition, Brazilian landowners that hold properties in Uruguay, managed to introduce slaves into the country as indentured labourers (*contratos de peonaje*). In 1862, new contracts of this kind were prohibited, but those signed before were still in force [18, 60].

[16] In 1889, of the total 59,000 employees in the city, 43,000 were foreigners. [13]

and entertainment [10,116]. At the same time, the thrust of modernism promoted the adoption of cultural models prescribed by the hegemonic class and in some cases the concealment of African origins [60]. But, the collective sense of Afro and Afro-descendants continued to be preserved in several ways, especially through social and civic organizations, called *sociedades* and *clubs*, and Carnival groups, called *comparsas*. They also constructed their own political clubs and black newspapers. In this process of identity formation, the traditionally Afro-Uruguayan neighbourhoods in Montevideo, namely Barrio Sur, Palermo and Cordón, played a prominent role; in particular, the tenement buildings called *conventillos* become culturally very significant and grew into core symbols of *candombe*, see Fig. 2.4.

Black societies, clubs and comparsas

During the second half of the 19th century, social, cultural and civic organizations proliferated in Montevideo, such as political clubs, artistic groups and mutual aid societies defined by European ethnicity. However, Afro-Uruguayans were banned from entering the society dances, social clubs, and other entertainment venues reserved for the middle and upper classes of the city.[17] Therefore, they created a racially segregated counterpart of this movement and founded their own parallel entities meant for the "coloured class" or the "coloured society" [18].

Towards the end of the 1860s, coexisting with *cofradías* and *nations*, part of the Afro-descendant population began to organize *societies* and *clubs*. For instance, the "Sociedad Pobres Negros Orientales" (Society of Poor Uruguayan Blacks) was founded in 1869, with the goal to create a music academy and to take part in Carnival, and the "Club Igualdad" (Club Equality) was created in 1872, with the main objective of organising a library [127].[18] The former also held monthly dances for members only, and twice a year open for all. This was an important service of these organizations to the Afro-Uruguayan community, because while the Africans remained devoted to their *candombes*, the new generation was fond of ballroom dances, and gave themselves up to the polka, the mazurka and the waltz [18].

To chronicle and promote all their activities, some of these organizations developed a fairly abundant press. Beginning with the foundation of the newspaper "La Conservación" in 1872, the black press served as a tool for spreading their political ideas and demands.[19] The struggle for equality of rights and opportuni-

[17] There were no direct racial bans, as this would have violated the country's guarantees of civic equality. However, references to de facto access restrictions are abundant, and become explicit in some cases, such as in a racial prohibition which appeared in an advertisement of Carnival dances in 1882, that was finally overturned by protests [18].

[18] There are references to many other organizations, such as the "Raza Africana" and "Negros Argentinos", which also took part in Carnival, and social clubs like "Club Progreso Social" and "Club Social 25 de Agosto" [127].

[19] Other Afro-Uruguayan newspapers from this period are: "El Progresista" (1873), "La Regeneración" (1884-1885), "El Periódico" (1889), "La Propaganda" (1893-1894, 1911-1912) and "La Verdad" (1911-1914) [127]. Between 1870 and 1950 Afro-Uruguayans produced at least twenty five newspapers aimed at black readers, the second-largest black press in Latin America in absolute terms, after Brazil, and by far the largest per capita [18].

ties and the unity of the black community are claims that appear recurrently in the editorials of the Afro-Uruguayan press [127]. By 1870, an ideological stance marked by positivism emerged from these editorials, that raised the rupture with the past that linked them with slavery. In this context, cultural practices related to the African were against the aspirations of "enlightenment" or "social progress" that they held. Years later, towards 1885, when the *nations* and *cofradías* were in their twilight, the editorials of the black press had an attitude of greater empathy with the past of their elders, but at the same time pointed out the differences in their life experiences as Afro-Uruguayans, for whom Africa as a homeland ceased to make sense. There was a clear dilemma regarding the possible forms of integration of the Afro-descendants into society, also expressed in the mix of cultural manifestations, which on the one hand reinterpret African legacy, and on the other hand, appropriate and transform the music and dance that was imposed through the European models of modernity [127].

To this respect, the charter of the "Sociedad Pobres Negros Orientales" is very instructive [224]. It states that its main goal is to create a music academy, that provides training in European musical instruments: piano, violin, flute and guitar. However, it also indicates that: "tambourines, castanets, drums, cymbals, triangles, and other African implements for the accompaniment of music are also understood to be instruments" [18]. Although not instruction on them was offered, it was assumed that the members would know how to play them and would use them in their performances, from what it seems that they considered African instruments and the music they made as a too rich cultural resource to be abandoned [18]. In addition, the charter stated that the *sociedad* would participate in Carnival as a *comparsa*, which shows the link between both types of organizations. This suggests they had a music repertoire in common, also shared with the dances they organized. In this repertoire—apart from the schottische, the polka, the mazurka and the waltz—they introduced a new form of music they called *tango*, resulting from the combination of elements of African and European origin [127].[20] It is argued that these early *tangos* borrowed heavily from the Cuban *habanera*,[21] which spread throughout the Atlantic coast—as music, song and dance—transmitted orally by trade between ports, and also through Spanish theatre and zarzuela, ballroom dancing and even art music of the time [127].

Since the colonial times, Africans and their descendants have participated in the popular festivities of the city with music and dance, Carnival being one of those occasions [9].[22] It was during Carnival that groups called *comparsas* paraded through the streets, singing songs and playing jokes on the passers-by [18]. The drive toward modernity of that time had major impacts on Carnival practices. Between 1867 and 1872 the number of such groups more than quadrupled, from 12 to 54, and their performance standards raised—improvised family *comparsas* left their place to groups governed by regulations, which included the role of a director [9, 18]. The members of these groups were dressed up representing

[20]By the 1920s, the *tangos* of the 1860-70s were described as "Creolized *candombe*" [89].
[21]Another Creolized musical form that combined African and European elements [127].
[22]The first documented reference to their participation in Carnival is from 1832 [222].

various carnivalesque identities, such as sailors, workers, laundresses, European immigrants, etc., among which the *comparsas de negros* were one of the most popular. The "Raza Africana" (African Race) was the first of these black *comparsas*, probably founded in 1867 [89], together with "Negros Argentinos" (Argentinian Blacks) and the above-mentioned "Pobres Negros Orientales", from the early period [127]. Over time, a large number of such black groups appeared,[23] becoming the vast majority of the existing *comparsas* by 1890 [127]. They also performed in dances and family houses, not only in Carnival but in other festivities [127].

Blackface comparsas

A different kind of *comparsa* burst into Carnival by 1876, the "Negros Lubolos", with the alleged goal to make known to the public the customs of the old negros of the African *nations* [18]. Surprisingly, its members were not Afro-Uruguayans, but young white man of middle- or upper-class background, that dyed themselves perfectly black with burnt cork and soot. They dressed in costumes evocative of those that were supposedly used by the members of the African *nations* and trained themselves in the type of songs and dances that blacks did at their *candombes* [18]. Like Afro-Uruguayan *comparsas*, they also sang and danced *tangos*, with a combination of European instruments and African drums [127]. Another blackface *comparsa*, the "Negros Esclavos", also appeared in scene in 1876, along with the "Negros Lubolos".[24] The blackface *comparsas* became very popular among all social classes and were featured at society balls [222]. Since then, *negro lubolo* became the term used in Montevideo's Carnival to designate a white man performing in blackface.[25]

The songs portrayed the enslaved and free Africans speaking heavily accented and grammatically incorrect Spanish, following the line of white poets and journalists of the time [18]. A nostalgic look back to Africa was one of the most recurring themes in their lyrics, as well as the yearning for white women's love, who were distant and untouchable for black men [18,127]. It is worth noting that Afro-Uruguayan *comparsas* addressed their heritage in different ways, some also exploiting the humorous representation of themselves, the comic effect of their amorous longing for white women forbidden by racial barriers, and remarking the black women's sexual appeal, who were, on the contrary, readily accessible [18,127,222].

The blackface *comparsa* is another case of a well-known phenomena of appropriation and reformulation of black musical forms by the white people [18]; like the blackface *minstresly* in the United States [195], or the blackface *teatro bufo* in

[23] Among them, "Nación Bayombe", "Negros Lucambas", "Nación Lucamba", "Negros Gramillas", "Negros Agunga", "Esclavos del Congo", and "Congos Humildes" [127]

[24] There are some references to previous participation of the "Negros Esclavos" in Carnival, between 1868 and 1870. The first blackface *comparsa* in Montevido's Carnival, "Los Negros", would have travelled from Buenos Aires to take part in it, from 1865 to 1867 [18].

[25] And it is remarkable to see in one of the songs by the "Negros Lubolos" from 1877, what may be the first appearance in print of the onomatopoeic representation of the *candombe* rhythm: *borocotó, chas chas*, that became widespread during the 20th century, arguably mainly among white population [18, 222].

Cuba [178], during the same time. For the white society, blackface was simultaneously a way to embrace and distance itself from blackness, while enabling the continued production and maintenance of racial difference and white supremacy. This dialectical counterpoint, involving fascination with and attraction to black culture, and at the same time marking the boundaries of class, race and gender, was unmistakably present in the songs of *negros lubolos* [18]. The lyrics paid respects to their white masters, ridiculed the pretentious black dandies aspiring to a higher social status, and portrayed black men as outsiders, unsuitable partners for romance or marriage [18]. Yet, they occasionally criticized racial inequality, for they were on firmer ground in attacking the established order, being young white men [18]. They also made use of the racial ventriloquism device [195], to dare to comment on the strict gender conventions and the puritanism prevailing at the time, which kept the middle- and upper-class young men and women apart at a safe distance [18, 33].

The appropriation and reformulation of the customs of the old negros was carried out, in part, by the creation of the stock characters of the *comparsa*, that still remain in the present-day Montevidean Carnival [18]. Two of these characters originated in the 1870s within the blackface *comparsas* [18, 89]. One of them is a reincarnation of the master of ceremonies of the African *nations*—the *bastonero*—who was then called the *escobero* (broomsman), since he used a broom as a symbol of command instead of his former baton (see Fig. 2.5). The other one is the *gramillero*, supposedly representing the traditional herbal doctor of the African *nations*, depicted as an elderly black man walking unsteadily on his cane, carrying a bag of herbs [222]. Years later, by the early 1900s, a new stock character was created, the *mama vieja* (old mother), an aged black women wearing a head cloth and a long skirt with petticoat, who carries a fan or an umbrella and dances with her partner, the *gramillero* (see Fig. 2.6). However, this new character was introduced by a different type of *comparsa* that developed during the late 1800s and into the 1900s, formed by working-class people, mainly European immigrants and their descendants, but which also practised racial integration [18].

Figure 2.5: Two *escoberos*, members of a *comparsa* in the beginning of the 20th century.

Proletarian comparsas and conventillos

In the 1860s and 1870s, when the black and blackface *comparsas* appeared, they were racially segregated, despite they were all generically called *sociedades/comparsas de negros*. However, by the first decade of the 1900s (and perhaps earlier) *comparsas* were no longer segregated, rather they included Afro- and Euro-Uruguayans, and European immigrants, together in the same groups. The level of ethnic integration varied largely across the different *comparsas*. For instance, the *comparsa* "Esclavos de Nyanza"—the most important of those years, founded in 1900—was almost entirely white and its members were mostly Spanish and Italian immigrants from the "La Facala" *conventillo* in the Palermo neighbourhood [18].

Large tenement houses, called *conventillos*, were erected in Montevideo since the 1860s, either from the reformulation of pre-existing residences or the construction of new buildings, intended as an accessible housing solution for low-income sectors of the society, namely rural population displaced to the city, European immigrants and people of African descent [3]. Thus, each family would rent a single room and share spaces such as bathrooms and kitchens. A large number of these *conventillos* were located in Barrio Sur, Palermo, and Cordón, the neighbourhoods that housed the greater part of the Afro-Uruguayan population at that time—and the *salas* of the African *nations*, during the second half of the 19th century.[26] From the cultural point of view, the *conventillos* of these neighbourhoods were always regarded as one of the most important social spaces in which people of African descent renewed and recreated their traditions, reinforcing a cultural identity that connected them with their origins [3, 9, 10, 13, 108]. Two important examples are the *conventillo* "Risso" (1885-1978)—called Mediomundo—in the Barrio Sur neighbourhood (see Fig. 2.7), and the *conventillo* "Barouquet" (1887-1965)—also known as Gaboto—in the Cordón neighbourhood. Another relevant case is that of the housing complex "Reus al Sur"—called Ansina— in the Palermo neighbourhood, which was not originally a *conventillo* but over time ended up functioning like one. See Fig. 2.4 for the location of these *conventillos*.[27]

It was in the *conventillos* that the immigrants came into direct contact with Afro-Uruguayan dances and music, and ultimately learned and reworked them [18]. Striking as it may sound, for them, one way to become Uruguayan was to take part in an African-based cultural form. As a consequence, a sort of working-class *negro lubolo* arose, and a new proletarian *comparsa* emerged, which integrated them together with the Afro-Uruguayans. Even in the majority-white *comparsas* of the early 1900, like the above mentioned "Esclavos de Nyanza",[28] at least a few black members were included, sometimes as directors of the group's drummers [18]. Some groups were more or less equally divided between whites and blacks, such as the "Congos Humildes" (1907) or the "Guerreros del Sur", and others were racially

[26] By 1867 there were already 115 *conventillos* only in Barrio Sur [3].

[27] The *conventillo* Mediomundo was located in 1080 Cuareim street, so it was also called Cuareim. The *conventillo* Gaboto was placed in 1665 Gaboto street. The Ansina housing was framed by the streets Lorenzo Carnelli, San Salvador, Minas and Isla de Flores. [13]

[28] Other majority-white groups were, "Libertadores de África", "Esclavos del Congo", "Esclavos de la Habana", "Esclavos de Asia" and reborn "Pobres Negros Orientales" [18].

mixed, but the existing records make it difficult to know in what proportions [18].[29]

The new proletarian *comparsas* adopted the existing stock characters of the *gramillero* and *escobero*, worked the same lyrical themes—including the melancholy for homeland in Africa—, and exploited the musical form of *candombe* or *tango* [18]. But they also innovated, not only by racial integration and the introduction of the *mama vieja*, but through an increased emphasis on drums and percussion instruments [18]. It was during the 1890s that the drum began to impose itself as the basic element of the *comparsa*, and then became its main musical instrument [224]. This clearly reversed the trend of the previous decades, which attempted to "civilize" the rhythms of the African *nations* with European instruments and melodies. According to the chronicles of the early 1900s, the drum corps continuously beating while marching through the streets produced a huge enthusiasm on neighbours and passers-by, and crowds followed the *comparsa* from the sidewalks while cheering and trying to imitate the drummers and dancers [18].

In this way, the primary template for the *comparsa* that paraded in Montevideo throughout the 1900s—and into the 21st century—was settled. The immense popularity of these groups and the drums' role in producing that popularity was also established. Besides, groups of drummers and dancers become ubiquitous in neighbourhoods such as Palermo and Barrio Sur, no longer confined to the days of Carnival. Some *comparsas* turned into identity symbols of the different neighbourhoods and developed very competitive, confronting each other when they met while parading through the streets, with dancing duels by the *escoberos* [18].

Even though until recently—in the 1850s and 1860s—*candombe* drumming was a heritage only circumscribed to the Africans and their descendants, it had now been disseminated through the working-class neighbourhoods and had been also adopted by European immigrants and Euro-Uruguayans. Arguably, in these neighbourhoods, drums continued to play much of the same role they played in the African *nations*, as a powerful tool for building community and social cohesion [18].

2.2.4 National rhythm: heritage and identity

During the first decades of the 20th century, the political and business elites began to devote efforts to turn Carnival into a commercial enterprise that would attract tourists from abroad. This increased the emphasis on staged theatrical performances and led to the emergence of new kinds of Carnival groups, such as *troupes* and *murgas*.[30] Everywhere in the city outdoor stages were erected—called *tablados*—where *comparsas* and other Carnival groups performed before the entire neighbourhood. Financially supported by the municipal authorities since 1923, its

[29] These include the "Pobres Negros Cubanos" (1890s), "Pobres Negros Hacheros" (1896), "Hijos de la Habana" (1912), "Guerreros de las Selvas Africanas" (1915), and "Libertadores de la Habana" (1915) [18].

[30] Developed during the 1910-20s, *murgas* are one of the main traditions of Montevidean Carnival. Consisting of a male chorus of about a dozen people, accompanied by three percussionists (bass drum, snare drum and a pair of crash cymbals), all make-up and wearing grotesque costumes, they sing satirical songs about politics and society. [8]

Figure 2.6: Traditional characters *mama vieja* and *gramillero*, members of *comparsa* "La Roma", parading in 2007 during the "Movida Joven" cultural festival (by Libertinus, on Flickr).

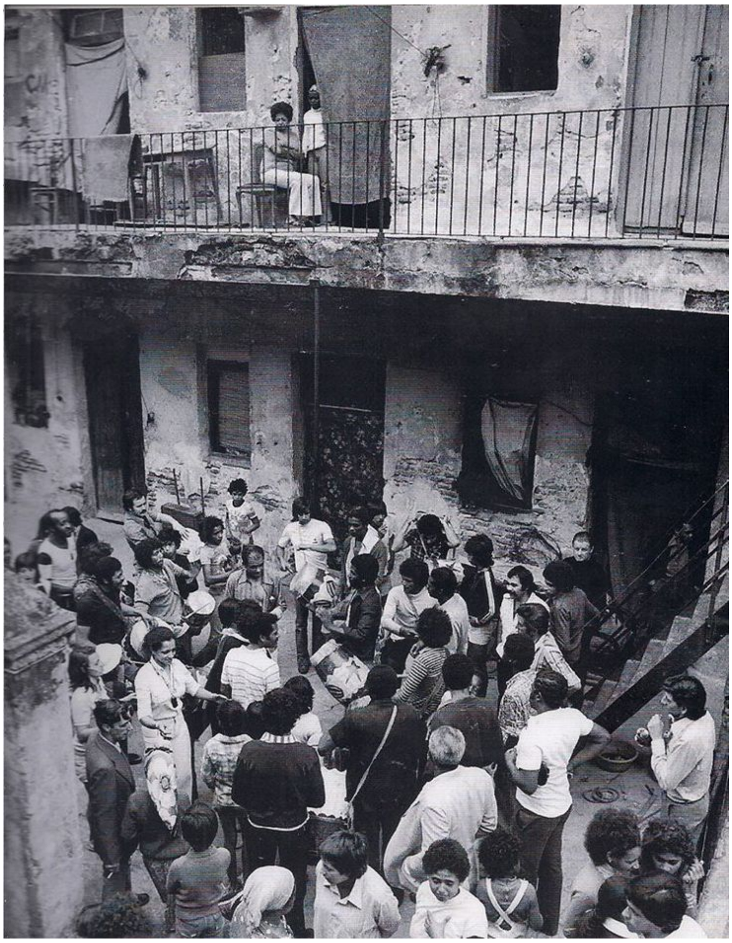


Figure 2.7: Conventillo Mediomundo. (Héctor Devia, from [13] reprinted with permission.)

number grew up to about 150 by the 1930s [18]. In 1944 a large open-air theatre called “Teatro de Verano” was inaugurated in the city, and since the following year, it has hosted the official contest of Carnival groups organized by the municipal government [11]. The increased expectations for more polished performances and the competition for municipal prizes influenced the profile of the *comparsas* [18].

Entering the Carnival contest in 1946, the group “Miscelánea Negra”, led by two white entrepreneurs, introduced several innovations to the traditional *comparsa*, to the point that the judges did not allow them to participate in the category of *sociedades de negros*. The novel music repertoire with more sophisticated musical arrangements, the showy costumes and the female dancers performing choreographic routines, were so welcomed by the public that judges decide to reward the *comparsa* with a special prize. Being then allowed to take part as a *comparsa*, “Miscelánea Negra”, won the first prize the following two years. In this way, the city government endorsed the importance of production values and entertaining show, promoting a new model for *comparsas*. Following this new tendency, another group by one of the same entrepreneurs, “Añoranzas Negras”, won the competition for the next five consecutive years (1949-53). Related to its success was the introduction of a new stock character, the *vedette*, that, instead of being connected to the Afro-Uruguayan past, was inspired from French cabaret.[31] Martha Gularte (real name Fermina Gularte Bautista), who featured as the *vedette* in “Añoranzas Negras”, was a great hit in 1949 and 1950, and began a career that ultimately lead her to become one of the greatest Carnival celebrities of all times [18]. Similarly, Rosa Luna, born in the Mediomundo *conventillo* in 1939, followed Gularte in the 1960s as another widely acclaimed *vedette* [13]. Since then, until the 1980s, all of the important *vedettes* were Afro-Uruguayan, as well as the overwhelming majority of the women of the corps of dancers. The stereotypes of black female sexuality and hot rhythm, both features supposedly carried in their blood, were reinforced by the introduction of the *vedette*. Therefore, despite the fact that it has turned into one of the most important characters of the *comparsas*, it is also controversial and some Afro-Uruguayans dissociate it from *candombe* traditions. Certainly, the supposed rhythmic and sexual ardency reflects a certain power associated with blackness, but it is not the kind of power that can produce social and economic progress and genuine racial equality, rather it can hinder upward mobility [18].

Despite Uruguay’s official doctrines of civic and social equality, discrimination of the “coloured persons” was all too common during the first half of 20th century. Reported mainly in the black press, it revealed the limits of Afro-Uruguayans integration into national life.[32] In the face of such treatment the Afro-Uruguayans continued to create their organizations, including Carnival *comparsas*, but also

[31] The first *vedette* was La Negra Johnson (real name Gloria Pérez Bravo), which actually paraded the year before. The role is probably influenced by Josephine Baker, the Afro-American performer that had been a sensation in Paris during the decades before. [18]

[32] Newspapers such as “La Vanguardia” (1928) and “Nuestra Raza” (1933) adopted a posture of racial militancy and regularly reported all sorts of racism and discrimination. But some notorious cases were dealt with by the press in general, like the controversy in 1956 surrounding Adelia Silva de Sosa, a young Afro-Uruguayan schoolteacher. [18]

newspapers, social and political clubs. Among them, the "Partido Autóctono Negro" (PAN, 1936), one of the three black political parties created in Latin America during the first half of the 1900s, and the "Asociación Cultural y Social Uruguay" (ACSU, 1941), which is still in existence today, making it one of the longest-lasting black social clubs anywhere in Latin America [18].[33]

It was ACSU that in 1955 devised a proposal submitted to the municipal authorities for an Afro-Uruguayan cultural festival, to take place during the twelve days of Christmas, between December 25 and January 6, commemorating the annual celebrations held in the past by the African *nations* (the coronation of the Kings and Queens of the Congo and Angola). The idea was welcomed by the authorities but as an opportunity to focus attention on Carnival. So the proposal was reconfigured into a special parade during Carnival, across Barrio Sur and Palermo neighbourhoods, called "Desfile Oficial de Llamadas", and devoted only to *comparsas*, that compete for cash prizes. The first *Llamadas* were held in 1956, and six groups took part, of which at least two were largely or entirely white. The first place of the competition was shared by two relatively new *comparsas*, "Fantasía Negra" (1954) and "Morenada" (1953), which though racially integrated, were directed by Afro-Uruguayans [18].[34]

Drawing its members from the Ansina housing complex in Palermo neighbourhood, "Fantasía Negra" was directed by Julio Giménez and Pedro Ferreira (real name Pedro Rafael Tabares, 1910–1980), who had worked together as directors of "Libertadores de África" during the 1940s. Being a composer and bandleader, Pedro Ferreira incorporated Afro-Cuban influences into *candombe* music—including rhythms, melodies and instrumentation—that he had learned first-hand by playing with Cuban rumba bands in Buenos Aires in the 1930s. Winning five first-place titles in Llamadas, from 1956 to 1963, and also five consecutive first-place titles in the Teatro de Verano (1954-58), "Fantasía Negra" become the dominant *comparsa* in the 1950s, up to the point of being asked by municipal authorities to depart from competition for several years to give others a chance [18].

On the other hand, "Morenada" was based in the Mediomundo *conventillo* in the Barrio Sur neighbourhood. It was directed by the Silva brothers, Juan Ángel, Raúl and Wellington, who after beginning their careers in "Añoranzas Negras" during the 1940s, formed their own drum corps called "Lonjas del Cuareim" and then founded a *comparsa* named "Morenada" in 1953 [13]. They won the first-place title in Carnival four times between 1959 and 1969, and five more times in the 1980s.[35]

The identification of the *comparsa* with the neighbourhood and the *conventillo* was very strong, and continued to be so in the following decades—not only Ansina

[33]Other examples are "Comité Pro-Homenaje a Don Manuel Antonio Ledesma, Ansina (1939), "Círculo de Intelectuales, Artistas, Periodistas y Escritores Negros" (1946), and "Movimiento Juvenil Independiente Pro Unidad de la Raza Negra" (1948) [18]. At present ACSU is known as "Asociación Cultural y Social Uruguay Negro" (ACSUN) [222].

[34]Recordings of the first Desfile de Llamadas in 1956 are available in CD [86], for three of the six groups, namely "Guerreros Africanos", "Morenada", and "La Candombera".

[35]While "Fantasía Negra" disbanded in the 1970s, "Morenada" was still active until the early 2000s, when the last of its founders, Juan Ángel Silva, died in 2003.

and Mediomundo, but also other in neighbourhoods such as the Gaboto *conventillo* in Cordón, with the *comparsas* "La Llamada de Gaboto" and "La Llamada del Cordón". Despite the fact that *comparsas* remained very competitive, there was also a lot of fraternity between them, which included the tradition of parading in feast days from one *conventillo* to another, to greet the adversary by showing off one's prowess and celebrate together [13]. Actually, the practice dates back to the times of the African *nations*, when its members would go drumming through the streets calling others with their distinctive rhythms, to gather in their *salas* or *sitios* and visit each other. The term *llamada de tambores* or simply *llamada* (drum call), refers to this tradition, in which groups of drummers and dancers, dressed in everyday clothes, parade along the neighbourhood in weekends and feast days [222].

During the 1970s and 1980s, some *comparsas* continued in the direction of increasing professionalism, innovation and spectacle. An Afro-Uruguayan dancer, choreographer and dressmaker, José de Lima (real name Carlos Lasalvia), was a key promoter of the kind of costumes and choreography he had witnessed in Rio de Janeiro Carnival while living in Brazil in the 1960s. Beginning in 1970, he created his own *comparsas*, "Serenata Africana" (1970–75, 1998–) and "Marabunta" (1976–93), to put his vision into practice [18]. The proposal was undoubtedly welcomed by the juries, considering the large number of first-place Carnival titles obtained by both *comparsas* in the 1970s and 1980s (five and seven, respectively). Another white dancer and choreographer, Julio Sosa (stage name Kanela or Canela), also followed this strand with his own *comparsas*, "Kanela y su Barakutanga" (1977–2001) and "Tronar de tambores" (2002–).

At the same time, other processes that strongly impacted the *candombe* culture took place in the 1970s and 1980s. During the civic-military dictatorship (1973–1985)—and also under the authoritarian governments of previous years—it became forbidden to meet in groups and to discuss collective issues publicly. In such circumstances, Carnival was subject to even closer regulation and control, and the spontaneous *llamadas*—calling people to drum and dance into the streets—acquired even more public significance [18]. Within a few years, the dictatorial government decided to demolished several of the historic *conventillos* and housing projects that had given rise to the traditional and best-known *comparsas*. In December 1978, the Mediomundo *conventillo* in Barrio Sur was evacuated.[36] The next month, in January 1979, residents of the Ansina housing project in Palermo received orders to vacate and most inhabitants were dislodged within the following few months. Finally, towards the beginning of 1981, Ansina was definitely deserted [13]. Other *conventillos* befell the same fate and hundreds of people were transferred to temporary accommodations, such as the sheds of an abandoned industry.[37] The government justified these actions on the poor conditions of the buildings and as a part of an urban renewal project. However, since the sites were

[36]In 2006, on the initiative of Edgardo Ortuño, one of the few Afro-Uruguayan parliamentarians in Uruguay's history, a law declared December 3 the annual National Day of *Candombe*, in tribute to the farewell *llamada* held in Mediomundo on that date in 1978.

[37]The Gaboto *conventillo* in Cordón was demolished some years before, in 1965.

not rebuilt and simply left as vacant lots, the desire to remove poor black and white families from the city centre and disperse them to outlying neighbourhoods was seen as the real motive. In addition, in later years black activists and organizations considered these forced evacuations as an assault to Afro-Uruguayan culture, which destroyed the birthplaces of the traditional *comparsas* [18].

Certainly, that was a shattering experience for people made homeless by the evacuations. And it was also difficult for the inhabitants that remained in the neighbourhoods, who were left without their meeting places and even with no partners to drum and dance [13]. However, it is also argued that if the government's goal was to undermine *candombe* it failed miserably [18]. The *comparsas* based on the demolished *conventillos* gradually returned to their activities. Former members of the Mediomundo continued to parade in "Morenada", and almost every year recalled their vanished home in their songs. Similarly, former members of the Ansina housing project then created "Concierto Lubolo" (1987-1997), as the lineal descendant of "Fantasía Negra", devoted to preserve and honour the memories of their gone birthplace. Years later, in 1999, Waldemar "Cachila" Silva—son of Juan Ángel Silva, founder of "Morenada"—created his own group and named it "C1080" in homage to Mediomundo's street address, Cuareim 1080. The following descendant from Ansina was the *comparsa* "Sinfonía de Ansina" (1994–2007, 2011–2014), directed by the Oviedo brothers, Gustavo and Edinson "Palo", who previously took part in "Concierto Lubolo". In this way, the lineage of traditional *comparsas* from these neighbourhoods continued to the present.[38]

During the 1990s and early 2000s there was a huge upsurge of *candombe* drumming groups, even in neighbourhoods not traditionally associated with them. The dispersion of the people from the *conventillos* throughout the city may have contributed to the phenomena, as well as the adoption of *candombe* drums by some highly popular artists in their recordings and performances.[39] Actually, whites were pouring in the *comparsas* in numbers never seen before, as if drums would have become fashionable [18]. Whereas the annual official Llamadas during the 1960s and 1970s had typically included six to eight *comparsas*, with corps of about 20 drummers,[40] by the end of the 1990s there were more than 30 groups performing, each of them formed by over one hundred drummers, dancers and performers. During the 2000s the parade had to be divided into two days, and an admission process was implemented to select only 40 groups, leaving many others out. The number of drummers and total members was limited to 70 and 150, respectively. The event has become massive, attended by tens of thousands of participants and beholders, and broadcast live.

[38] The *comparsa* "C1080" is still active and won several first-place titles in Carnival and Llamadas. An actual *comparsa* from the Ansina tradition is "Valores de Ansina" (2014–).

[39] In 1984, Jaime Roos recorded in studio two songs with *candombe* drums, namely "Tal vez Cheché" and "Pirucho", featuring Gustavo and Edinson Oviedo, Fernando 'Hurón' Silva and Fernando 'Lobo' Nuñez [12]. By the same year, Alfredo Zitarrosa recorded "Candombe del olvido" in studio, and José Carbajal recorded "Ya comienza" live, both using *candombe* drums [237]. The latter also features the Oviedo brothers and Silva.

[40] See for instance the CD [86], for the number of *comparsas* recorded in the Llamadas in 1960, 1962 and 1964, and also the number of drummers reported in 1964 for each group.

The spontaneous *llamadas* of “Candombe de Reyes”, held on January 6th in Barrio Sur and Palermo—which follow a tradition since the colonial period—has involved increasing amounts of participants and spectators in the last two decades, assimilating more elements of spectacle and Carnival, to the detriment of the less produced and informal groups of the neighbourhoods [126, 236]. In addition, many more *candombe* drumming groups play informally in the neighbours all over Montevideo. Since the 1990s, official parades organized by local authorities have also emerged in other cities of the country, such as Durazno and Canelones, which bring together thousands of people and dozens of *comparsas* [237, 260].

The widespread growth of *candombe* brings about some positive aspects, such as greater social and institutional legitimacy, but also raises some problematic issues. From a musical perspective, master drummers always stress the fact that *candombe* drumming is essentially a dialogue, with call and response interactions.[41] But, they argue, most new drummers do not have enough *candombe* vocabulary to listen to and understand these dialogues, let alone to take part in them. Besides, the large number of members of the modern drum corps makes it difficult to hear and respond to the conversations [18]. The traditional process of gradually learning each drum at a time, to master one before attempting the other (first *chico*, then *piano* and finally *repique*), has been subverted by an alleged tendency to immediacy. At the same time, certain practices that require a hierarchical organization—such as the end of the *llamada*, in which all the drums stop playing precisely at the same time, or the control of tempo variations—are more troublesome to implement. In face of such difficulties, some master performers began offering workshops on *candombe* drumming—only learnt by imitation in the past—, while others simply retired from Carnival.[42] The performing styles were also affected by the new situation. In the past, each neighbourhood had a distinctive and recognizable style of performing the rhythm, the three most important being Barrio Sur (Cuareim), Palermo (Ansina) and Cordón (Gaboto) [5].[43] But with the emergence of so many new groups, styles became increasingly diffuse and the knowledge of their former differences is scarce, even among those who act as juries in Carnival [88]. In addition, the spiritual aspects attributed to the *llamada de tambores* [106, 111], and even to the official Llamadas parade as a memorial ritual [22], are unknown to the majority of those who recently approached *candombe*.

Another frequently mentioned issue of concern is that, despite the commercial exploitation of Carnival and the Llamadas, which supposedly generates significant revenues, it provides a modest living for a handful of stars, but it is particularly unrewarding, in financial terms, for most of those who take part. The poverty in

[41] For the following assertions, see the interviews with Tomás Olivera Chirimini, Sergio Ortuño, and the brothers Héctor Noé and Fernando Nuñez (son), published in DVD [88].

[42] In 1996, Sergio Ortuño with Miguel García founded a *candombe* school in Mundo Afro, today lead by Álvaro Salas Gularte. Since 2006, Sergio Ortuño has carried out his own school within the Triangulación-Kultural project. Master drummer Héctor Manuel Suárez founded in 2002 a *candombe* school called Integración, which is still active.

[43] Recordings of Barrio Sur and Cordón styles in the 1960s are available in CD [86]. Modern examples of the three styles, both in small ensembles and *comparsas*, are available in CD [210] and DVDs [87, 88].

which a *comparsa* director or a renowned master drummer lives—some of which not even own their own drums—can be striking [18]. This is consistent with the broader picture of ethnic-racial inequality in the country, since the poverty rates of the Afro-Uruguayan population more than double those of the rest of the society [61, 65].[44] According to recent studies, Afro-descendants present a clearly unfavourable situation in all indicators related to educational and economic performance. They have lower average income levels, reach lower levels of education, and generally work in low-skilled and lower paying jobs. At the same time, they have little participation in positions of direction, in the politics and in the academy [64, 65].

The surveyed data on racial inequality, available since 1998, has supported the demands of black activists and organizations, such as Mundo Afro,[45] for government efforts to assist black population. Only recently, based on two laws from 2006 and 2013, specific initiatives and affirmative actions are being gradually implemented.[46]

In 2009, in a context of legitimation and appropriation by broad sectors of the society, *candombe* was inscribed on the Representative List of the Intangible Cultural Heritage of Humanity by UNESCO, being considered a symbol of the identity of communities of African descent in Montevideo. This constitutes the greatest institutional recognition that a cultural practice can obtain. As well, it entails responsibilities for the political classes and the society at large, in particular on ensuring its continuity. This implies maintaining its essential meanings and elements of social cohesion, as well as guaranteeing the welfare of those who have developed them. The academia is also urged to contribute to its preservation, through research and outreach, and its participation in political decision-making [236]. From the "Ministerio de Educación y Cultura" (MEC, Ministry of Education and Culture), some initiatives have been taken, such as the creation of the "Grupo Asesor del Candombe" (Candombe Advisory Group) formed by referents from the Barrio Sur, Palermo and Cordón, and the implementation of various actions for the elaboration of a preservation plan [237, 260]. Also within the MEC, the "Centro de Documentación Musical, Lauro Ayestarán" (CDM) has organized several academic and outreach events, and has edited valuable bibliographic and audiovisual material, including both historical records and new productions [86–88, 112, 158, 236].

What in the past was a cultural expression confined only to the descendants of Africans, today is the heritage of all Uruguayans, and even of the whole human race. Thus, the typifying and the descriptive use of the Afro prefix for *candombe* is imprecise, as it contains both a false generalization and a false reduction [236]. In

[44] In 2006, 50 percent of the Afro-Uruguayans fell bellow the national poverty line, versus 24 percent of whites [61]. In 2012, the incidence of poverty among the Afro-descendants was 27.2 percent, while in the total population it was 12.4 percent [65].

[45] Founded in 1998, by a group of former members of ACSU led by Romero Rodríguez and Beatriz Ramírez, it has been by far the most visible of the Afro-Uruguayan social and civic organizations during the 1990s and early 2000s. It gathers different subgroups with specific interests and missions. [18,256] Other present organizations are Casa de la Cultura Afrouruguaya, Coordinadora Nacional Afro, Mizangas and Triangulación Kultural.

[46] Law number 18.059, "Candombe, Cultura Afrouruguaya y Equidad Racial" (2006), and law number 19.122, "Acciones Afirmativas para Afrodescendientes" (2013). [247]

this regard, as remarked by Alfaro [13], on the one hand, one can note the racist component that lies beneath the discourse that confines *candombe* only to blacks. On the other hand, it can be condemned as being racist the supposed frivolity with which some whites would claim *candombe* as their own. Both positions refer to a key controversy, which has to do with the losses and gains that are part of cultural change, which allows a tradition to stay alive and continue interacting with people [13]. What seems to be crystal clear is that, for part of the black population *candombe* is their most precious patrimony, and far from being a fashion, plays a central role in their culture and their vision of the world [13]. Precisely in a part of the world where blackness has more than often been historically and culturally assumed to be "invisible", as Andrews points out [17], the *candombe* drums project in the modern urban space powerful sound memories of Uruguay's past and its links to Africa. Although controversial, complicated, problematic, and also biased as can be such memories, they are much better than silence [17].

2.3 Candombe drumming

Although originated in Uruguay, the practice of *candombe* discloses its strong African roots in its instruments' topology, its rhythm, and its performance practices. The core element of this tradition is the *candombe* drumming, performed by groups of drummers playing a distinctive rhythm while marching on the street. This section is devoted to describing the *candombe* drums, the essential patterns of the rhythm and its resulting metric structure. Finally, a brief account of the main influences of *candombe* drumming into the popular music of Uruguay is given.

2.3.1 Candombe drums

The instrument used in *candombe* is simply called *tambor*,[47] the generic Spanish word for drum, of which there are three different sizes with their respective registers: *chico*, the smallest and high pitched, *repique*, of medium size and register, and *piano*, the biggest and low pitched.[48] An ensemble of drums is called *cuerda*, which in its minimal form consists of one of each type, but can gather dozens of drums. Fig. 2.8 shows, from left to right, a *chico*, a *repique*, and a *piano*.

The drumhead is hit with one hand bare and the other holding a stick, that is also used to hit the wooden shell of the drum to produce a sound called *madera*. The rhythm is played while marching, so the drum is carried hanging from the performer's shoulder by using a strap, called *talí* or *talín* [106]. An animal hide is used for the drumhead—usually cow's hide, and sometimes colt's skin—, whereas

[47]The denomination *tamboril* was also widely used in the past and can still be found in some cases. But currently the performers prefer the term *tambor*. Some scholars point out that *tamboril* allows for certain specificity compared to the generic word *tambor* [5].

[48]The *chico* drum was also called *pino* by the early 1950s, and the terms *congo*, *pelé* or *belé* were used in the past too. An old name for the *piano* drum was Gon or Ngongon [106].

the bottom end is open.[49] The thickness of the drumhead varies for the different drums: thinner for *chico* and *repique*, and thicker for *piano*. Traditionally the skin is nailed to the shell of the drum. Thus, the tuning of the drum is achieved by heating the skin by means of a fire [106], usually prepared in the street next to the kerbstone, as shown in Fig. 2.9. This process, which is called *templado* (literally warm-up), constitutes one of the most characteristic rites of *candombe* drumming and a privileged moment for social interaction. Most modern drums, however, use a mechanic tuning system in order to tighten the drumhead (see Fig. 2.8). The skin is mounted on a metal hoop, and by using a wrench, tension rods are screwed into threaded lugs attached to the drum's body. Yet, even drums with mechanic tuning systems are often tempered to the fire as well (see Fig. 2.9). The tuning is approximate, but has to provide the relative pitch between the three different type of drums, such that each one lies in its corresponding frequency register [5, 106].

Figure 2.8: Candombe drums, from left to right: *chico*, *repique* and *piano*. (*Martín Rocamora*)

In the past, the drums were built from barrels—mainly barrels of "yerba mate"—the hogshead staves modified to build drums of different sizes [31]. By the end of the 1960s, the barrels were no longer in use and become scarce, so the drums began to be made from scratch. Different kinds of wood have been used for that purpose, such as pine, oak and cedar—the former widely preferred nowadays for being lighter. The changes in construction methods also impacted the shape and dimensions of the drums. The drums in the past were straighter than their modern version [5]. According to drum builder Juan Velorio (real name Bienvenido Martínez), by 1959 the size of the drums was increased.[50] The height of all of them is about 70 to 80 cm, but it is the diameter and the amount of bent that varies for the different drum types. For instance, Juan Velorio gives the

[49]During the 1980–90s, plastic drumheads were also used, manufactured for other percussion instruments or of disused radiographic film. Currently they are no longer in use.

[50]See the interview "Juan Velorio, ingeniero del tambor", ca. 2000, TV Ciudad. He also states that the size increase led to the disappearance of the *bombo* drum, which was a *piano*-like drum but of a larger size. The shape and dimensions of a group of drums in 1965 is reported by Ayestarán in [31], also including a *bombo*, referred to as *bajo*.

following values for the diameter of the drumhead: 10 and 11 inches for *chico*, 12, 13 and 14 inches for *repique*, and 15 inches and beyond for *piano*. Nonetheless, other drum builders may have different criteria, for example, some advocate for a smaller *chico*, of about 7 inches, which yields a higher pitch.[51]

Figure 2.9: *Templado de tambores* before the Desfile de Llamadas 2005, *comparsa* "C1080". Note that the *piano* drum in the middle has no mechanic tuning system. (*Martín Rocamora*)

The construction of drums is done in an artisan way, as shown in Fig. 2.10, with no manufacturers companies involved, and the trade is passed from masters to apprentices, following a sort of lineage. For instance, Juan Velorio learned the trade with 'Quico' Acosta, after working several years for him disarming barrels.[52] Likewise, Fernando 'Lobo' Núñez, learned with 'Cabeza' Montrasi, who in turn had learned with Juan Velorio. Currently, Nuñez passed the trade to his sons Héctor Noé and Fernando Nuñez, while 'Catito' Martínez, son of Juan Velorio, is also building drums.[53]

2.3.2 Rhythmic patterns and metrical structure

The analysis of various corpora of African and Afro-diasporic music reveals a musical organization generally composed of recurrent temporal patterns [4, 15, 20, 39, 119, 177, 190, 217]. This is also the case of *candombe* drumming, whose rhythm results from the interaction of rhythmic patterns of the three different drums.

It is important to recall here that *meter* and *rhythm* are two different but interrelated concepts [193]. The area of rhythm has deserved a lot of attention in recent music theory, leading to the development of a new conception of meter in the past few decades. One of the most important contributions to this new view of meter was Lerdahl and Jackendoff's Generative Theory of Tonal Music

[51] See the interview with Héctor Noé and Fernando Nuñez (son), published in DVD [88].

[52] By 1966, Ayestarán conducted two interviews with drum builders Eulogio 'Gitano' Celestino and Valentín 'Gaucho' Piñeyro, that provide valuable information on the construction process. It seems that Piñeyro also learned the trade from 'Quico' Acosta [260].

[53] In the 1990s, *candombe* drums by the late Alfredo 'Pocho' Guillerón, began to be sold in music stores. At present, Álvaro Rabasquiño is another accomplished drum builder.

Figure 2.10: Fernando 'Lobo' Nuñez building drums in his workshop in Barrio Sur.

(GTTM) [181]. The metrical structure is regarded as a regular pattern of points in time, called *beats*, hierarchically organized in levels. However, beats do not necessarily correspond to the actual events present in the piece. Rather, the listener must infer the metrical structure from the events of the piece. What is more, once the meter is established, the events of the piece need not constantly reinforce it, and may even conflict with it to some extent [193, 281].

Conversely, the concept of rhythm involves patterns of organised durations that are phenomenally present in the music [169, 193], i.e. they form the music stimulus itself, also called 'surface rhythm' [142]. While meter is a mental construct not directly heard but inferred from the rhythm, its organised pulsation provides specific anchor points to structure the rhythm [75]. Hence, meter is also regarded as a background to the rhythmic pattern, which is based on the Kolinski's idea of meter as a framework [169]. It has been noted that rhythms in African and Afro-diasporic music often exhibit a significant portion of note onsets not directly reinforcing the metrical structure, by allowing anti-phase (off-beat) and other more complex types of non-congruent relationships between rhythm and meter [62, 169, 190, 194, 217]. However, there is currently a strong agreement on that the metric frameworks underlying these rhythms are usually the same found in Western music [119, 190, 194, 281]. The most common meters involve a cycle of four even beats that are subdivided into two, three, or four faster pulses [21, 62, 190, 194, 238].

In *candombe* drumming each drum has a distinctive rhythmic pattern, associated to its specific register, and corresponding to a different musical role. The *chico* drum functions as a timekeeper, playing an ostinato in the high register, which defines the lowest metrical level (sometimes called 'density referent' in African musicology [217]). The *repique* plays in the mid register and acts as the variative lead drum, whereas the *piano* drum plays in the low register an accompaniment that allows for different variations and ornamentations. Besides, the rhythm has a timeline pattern,[54] called *madera* (or in analogy to Afro–Cuban music, *clave*), which is shared by the three drums and has many traits in common with similar patterns found in Afro–American music, like the *son* clave. It is also worth noting that the drummers move forward walking with short steps synchronized with the beats or *tactus*, and the transference of the body weight from leg to leg while marching constitutes a fundamental pattern, not audible but internally felt [110].

The rhythmic pattern of the *chico* drum is depicted in Fig. 2.11, where lower and upper line represent hand and stick strokes respectively. Following a virtually immutable pattern, the *chico* drum defines the *tatum*, i.e. the lowest level of pulsation over which the metric structure is built. This basic pulse is usually played at a high rate, typically from 450 to 600 beats per minute (beats per minute, the unit used to measure tempo (bpm)). The periodicity of order four of the pattern is in the range of about 110 to 150 bpm and is perceived as the *tactus*. However, the location of the beat within the pattern can be very difficult to perceive without any further references, being the hand stroke—which is accented—the strongest candidate [158].

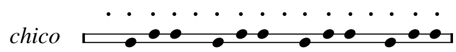


Figure 2.11: Rhythmic pattern of the *chico* drum. Lower and upper line represent hand and stick strokes respectively, a convention that is used henceforth in all music notations.

The *clave* pattern is produced by hitting the shell of the drum with the stick, and is played by all the drums as an introduction to and preparation of the *llamada* rhythm; then it is also played by the *repique* drum in between phrases (see the first *repique* on the left of Fig. 2.1). As in other Afro-Atlantic music traditions, the *clave* serves as a mean of temporal organization and synchronization. Fig. 2.12 shows the kernel of the rhythmic structure of *candombe*: the superposition of the *chico* pattern and the *clave*. The role of the *clave* is twofold: it establishes the location of the beat with respect to the *chico* pattern, and also defines a cycle of four beats (sixteen *tatum* pulses), thus inducing a higher metrical level [158].[55] The way the *madera* pattern operates in the *candombe* rhythm, however, presents interesting differences with the more common uses of timeline patterns in other Afro-Atlantic music traditions. For instance, instead of a single timeline pattern

[54]The concept of 'timeline' is essential in understanding and analysing the rhythmic organization of the music of Africa and the Afro–Atlantic diaspora. Arthur M. Jones was the first to underline the importance of the bell pattern in the music of the Ewe people of West Africa and its pervasiveness in sub–Saharan music [153, 154]. Years later, Kwabena Nketia introduced the term timeline, now widely used to refer to a short rhythmic pattern repeated cyclically in the manner of an ostinato [217]. This pattern serves as a reference for temporal organization and as an identifier for each rhythm or "song". Among several others, terms like timeline, bell pattern, standard pattern, guideline or, in Latin America, *clave*, have been used to refer to this key element of the rhythmic structure [160].

[55]The first beat of this rhythm cycle is called the *downbeat* in this dissertation.

as in other Afro–Latin–American musics, the clave pattern in *candombe* allows for several different types and variants [160].

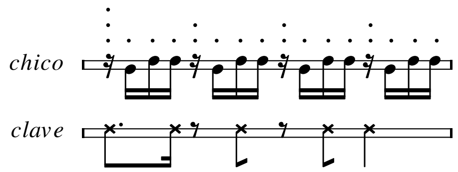


Figure 2.12: Interaction of *chico* and *clave* patterns, and the three levels of the resulting metric structure. There are actually several possible variations of the *clave* pattern shown here. For example, the third and fourth strokes can be displaced, and/or additional strokes can be added.

Two characteristic traits link this rhythmic/metric configuration with other Afro-Atlantic music traditions: 1) the pattern defining the pulse does not articulate the *tatum* that falls on the beat, and has instead a strong accent on the second; 2) the *clave* divides the 16-*tatum* cycle irregularly (3+3+4+2+4), with only two of its five strokes coinciding with the beat. In this respect, the rhythmic/metric structure of *candombe* differs from tonal metric structures found in Western music [181], making it difficult to decode for listeners not familiar with it [219].

The *repique* and *piano* drums are both technically and rhythmically much more complex than the *chico* drum, exhibiting also more variation in their playing. Fig. 2.13 shows their respective patterns simplified to their essentials, along with the metric structure and the previously introduced *chico* and *clave* patterns. The *repique* has the greatest degree of freedom among the three drums: by exploiting a wide repertoire of complex variations in its rhythmic patterns, it is the main responsible for generating interest, surprise and musical variety in *candombe*. It has, however, a primary pattern, shown here in its basic form [157, 158].

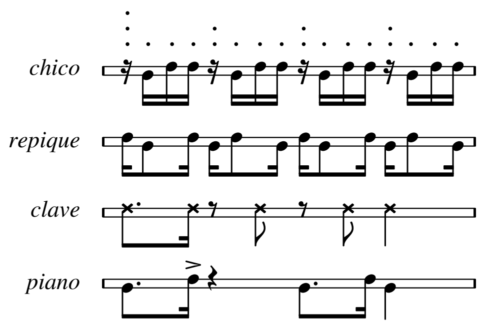


Figure 2.13: Interaction of the main *candombe* patterns and the resulting metric structure. The patterns of the *repique* and *piano* drums are shown in a simplified basic form.

With respect to the *piano* drum, it can be seen that, reduced to its rhythmic skeleton, its pattern is congruent with the *clave*, thus reinforcing the four-beat cycle defined by the latter [158].[56] Therefore, its musical role can be assimilated

[56]It has been note the similarity of the essentials of the *piano* drum pattern to the *habanera* pattern [111,127]. Actually, various rhythmic structures are shared among closely

to a timeline pattern. Yet, the *piano* drum actually has two functions: playing the base rhythm (*piano base*), and occasional more complex figurations (*piano repicado*). These can be either ornamented variations of the *base* pattern, or figurations derived from the primary pattern of the *repique* drum (hence the name). Some notable *piano* players developed distinctive rhythmic patterns and personal performing styles that influenced other players. Besides, it is in the patterns of the *piano* drum that the differences among the styles of the different neighbourhoods (*barrios*) are more evident.

The three more important traditional styles, from which all others derived, are Barrio Sur (Cuareim), Palermo (Ansina) and Cordón (Gaboto) [5]. Indeed, the two more different styles are Barrio Sur and Palermo [106], while Cordón style is often regarded as a variant of Palermo. A *piano* performance in the Barrio Sur style typically includes less *repicado* patterns, so that the *base* patterns—often embellished—prevail, whereas the Palermo style is characterised by frequent *repicado* patterns that produce call-and-response interactions between *piano* and *repique* drums. In addition to the distinction of the *piano* drums, other often mentioned stylistic difference concerns the *tempo* of the performances. The Barrio Sur style usually begins the rhythm slowly, and despite the fact that tempo can be increased throughout the performance, low tempos are typically maintained for long periods. On the other hand, the Palermo style tends to be faster, and even though it can also exhibit tempo variations, there is usually an intention of tempo increase but only up to a certain value in order to keep the rhythm comfortable for playing and dancing. In the style of Cordón the rhythm is played at an even higher tempo.[57]

2.3.3 Candombe drumming and popular music

In addition to its early links to *tango* by the late 1800s [127], *candombe* drumming has influenced several kinds of popular music in Uruguay throughout the 20th century [5, 221, 235]. During the first decades of the 1900s it was mainly restricted to the *comparsas* in Carnival, and to the spontaneous *llamadas* through the historic black neighbourhoods of Montevideo in feast days or public holidays. The drums were also used in private family gatherings of the Afro-Uruguayans, during which not only *candombe* was played, but also other rhythms such as the *habanera*, the so-called *milongón* (a slow-tempo *candombe*) and some ternary rhythms (6/8, 12/8) generically called *afros*.[58] In fact, all these rhythms were part of the repertoire of

---

related musical forms such as *tango*, *candombe*, *milonga* and *habanera* [127]. For instance, the *habanera* pattern and its variants usually appear in the low frequency register of these music styles, whereas the group of 'headless' sixteenth notes characteristic of the *chico* drum, and the sixteenth–eighth–sixteenth note motif of the *repique* primary pattern (as well as the eighth-note triplet) are found in the mid and high frequencies [127].

[57] For style comparisons, see the interviews with Tomás Olivera Chirimini, Sergio Ortuño, Wellington Suárez, and the brothers Héctor Noé and Fernando Nuñez (son), published in DVD [88]. Recordings of the three styles are available in CD [210] and DVDs [87, 88].

[58] The first description of this practice is due to Ayestarán by 1967, who named it "conversación de tambores" (drums conversation) [29]. He noted that the performers

a *comparsa*'s stage performance, and are used at present in the same context.

Since the mid 1930s, the orchestras of *tango rioplatense* introduced *candombe* influences, labelled by different names, such as *tango candombe* or *milonga tango* [237]. Among the pioneering composers were Alberto Mastra (real name Alberto Mastrascusa, 1909–1976) and Horacio "Pintín" Castellanos (1905–1983) [221]. The noted Afro-Uruguayan *candombe* singer Lágrima Ríos (real name Lida Melba Benavídez, 1924–2006), also stood out in this genre. The movement was consolidated by the 1940s, with some recordings including *candombe* drums, but would be extinguished a few years later, before the 1960s [237].

A different musical strand was developed since the mid 1950s by the renowned Afro-Uruguayan musician Pedro Ferreira (real name Pedro Rafael Tabares, 1910–1980). By the time he was director of *comparsa* "Fantasía Negra", he created an orchestra named "Cubanacán", that introduced Afro-Cuban influences into *candombe* music, mainly from the successful "Lecuona Cuban Boys" orchestra. This music become a stylistic reference for *comparsas* and other *candombe* groups since then, and some of his compositions turned iconic [237].

By 1964, singer and producer George Roos (1925–1995), promoted the fusion of *candombe* and jazz, a project known as "candombe de vanguardia" (avant-garde *candombe*), with the idea of exploiting it as a dance form for large audiences [221]. Three different groups were created, drawing some of its members from a jazz club in Montevideo, and a record was produced for each of them.[59] Despite the lack of commercial impact and the very few shows they had, some of their songs become *candombe* 'standards' and inspired the following generations of musicians [83].

Also by the mid 1960s, a movement of folkloristic popular song—known as *canto popular*—was developed, in opposition to the authoritarianism of the government and the subsequent dictatorship, reaching very large audiences [237]. Some of these artists included *candombe* influences in their music, such as Alfredo Zitarrosa (1936–1989), José Carbajal (1943–2010) and the duet "Los Olimareños", formed by Braulio López (1942–) and José Luis "Pepe" Guerra (1943–).

In the late 1960s, influenced by rock—particularly The Beatles—a younger generation of musicians created a movement of popular music known as *candombe beat*. The first group of this kind was "El Kinto" [83], founded in 1967 and active until the mid 1970s, which was lead by two very influential musicians, namely Eduardo Mateo (1940–1990) and Rubén Rada (1943–). In 1968, an album was edited by the rock group "Los Shakers", lead by brothers Hugo Fattoruso (1943–) and Osvaldo Fattoruso (1948–2012), including the song "Candombe", the first one to be released in this style [83]. By the early 1970s Rada lead the group "Totem", and years later, after following his solo career, he would become the most renowned

were sitting, holding the drums between their legs and playing with two bare hands. The performances are described as highly improvised and refined, involving several rhythms.

[59]The directors of the groups were Manuel "Manolo" Guardia (1938–2013), Hebert Escayola (1928–), and Daniel "Bachicha" Lencina (1938–2017), while the singers were the talented Afro-Uruguayans Cachito Bembé (real name Fermín Adolfo Ramos) and Hugo "Cheché" Santos (1941–), both with fruitful careers in Carnival groups. The recordings featured *candombe* drums along with different jazz ensembles.

artist of the Afro-Uruguayan music, and one of the most important in Uruguay, reaching massive success [245], see Fig. 2.14.

By the mid 1980s, a process of greater promotion of *candombe* takes place, in the context of its revaluation as an element of identity within the popular music movement [237]. A sign of that is the adoption of *candombe* drums by very popular artists like Jaime Roos (1953–), Alfredo Zitarrosa, and José Carbajal, in their live performances and recordings [12, 237]. Previously, the influential group "Opa", led by the Fattoruso brothers, had included a *cuerda de tambores* in their concert in 1981. Actually, it seems that it was the Afro-Uruguayan musician Jorginho Gularte (1956–2013, son of Martha Gularte) who, being invited to take part in that show, proposed the inclusion of *candombe* drums [237]. Then, since the 1990s, musicians from the popular music movement have been participating in *comparsas* during Carnival, as is the case of the *comparsa* "Sarabanda" in 1992 [237]. Currently there are groups exclusively devoted to *candombe* music, such as "Conjunto Bantú" (1971), "La Calenda Beat" (1982), and "Rey Tambor" (2007), or artists primarily linked to it, like the Afro-Uruguayans Eduardo Da Luz (1954–) and Isabel "Chabela" Ramírez (1958–). While *candombe* rhythm has been adapted to the drum set and other percussion instruments like congas,[60] it is common to find bands that include a small *cuerda de tambores* as rhythmic accompaniment, see Fig. 2.14.

Figure 2.14: Rubén Rada at a live show with *candombe* drums as rhythmic accompaniment.

[60]Since the 1970s there has been a very respectful approach by some non-Afro-Uruguayan musicians [6], mainly percussionists, which led to methodical reformulations of *candombe* rhythm to the drum set and percussion instruments, see for instance [257].

Chapter 3

Data collection and generation

This chapter is based on work originally reported in [255] and [219]. The description herein reproduces some passages of the articles and also includes modifications and additions in order to put the work in the context of this dissertation.

3.1 Introduction

The development of computational tools for musicological studies and musical analysis has been a very active and promising area of research in recent years [291]. It is always essential for this type of research to count on a representative annotated dataset of the corpus under study, both for the development and testing of the techniques and tools, and for the proper musicological analysis. Several datasets for development and evaluation of MIR applications have been released, from the pioneering and widely used RWC database [131–133] to more recent endeavours such as the Million Songs Database [43] or the MendelyDB [47]. While most of them are devoted to audio recordings, the ENST-Drums database [121] is a collection of audio and video recordings for drum signal processing. The availability of suitable annotations—such as beat structure, chords or melody line—is a critical issue that determines the usefulness of such datasets [182]. Even though some software tools have been developed to alleviate the process [66], producing those labels is usually a very time consuming task, manually done by music experts.

The present chapter is devoted to the description of the data and music collections used during this thesis work. The musical setting considered is small-size *candombe* ensembles, of three to five drums. Since the sort of data needed to conduct this type of research were not previously available, an important amount of work was dedicated to collecting and labelling audio recordings. In addition, some software tools were devised to render synthetic rhythmic patterns into audio files in order to carry out some studies. Besides, recording sessions were also conducted with the aim to produce an audio-visual database of *candombe* performances, that could serve both documentary and research purposes.

The following section describes the generation of synthetic rhythmic patterns and performances, together with ground-truth annotations. After that, Section 3.3

presents a dataset of labelled *candombe* recordings for beat and downbeat tracking that was produced during this thesis work and is available for the research community. In Section 3.4, the process of creating the audio-visual database of *candombe* performances is detailed, including the annotation efforts carried out and some of the technical challenges tackled.

3.2 Synthetic rhythmic patterns and performances

During the development of the techniques proposed in this work, it turned out that testing them under tightly controlled situations was extremely useful. For this reason, a set of software tools was devised to produce sample-based synthesized *candombe* patterns. This avoids some aspects concerning performance, such as micro-timing variations, or regarding recording conditions, such as reverberation and noise, and simplifies the process of creating the ground-truth labels. The ability to precisely synthesize audio files from scores proved to be a very valuable research resource. For instance, the same synthetic example rendered at different beats per minute allowed a detail study of the influence of tempo in the performance of the beat tracking algorithms. The tools were crafted by Luis Jure, and the author of this thesis put them into practice and implemented some features, such as the generation of onset labels.

The process for creating a synthetic *candombe* pattern encompasses the following steps. Firstly, a score is produced using the `Lilypond`[1] music engraving software language, adopting some conventions to represent the different instruments and types of strokes (as introduced in Section 2.3.2). By compiling this code, a sheet music file is obtained, along with a text file containing a list of events and control messages. Then, this information is parsed and processed using a `Python` program to produce a score file that is interpreted by an "orchestra" written in the `Csound`[2] sound synthesis software. Several samples of each type of stroke, previously recorded by a professional musician, are selected randomly by the synthesis program, which is able to interpret local accents and gradual variations in dynamics (e.g. *crescendo*), as well as tempo indications and progressive changes (e.g. *accelerando*) from the score. According to the onset time, the type of sound, and some additional information such as dynamics, an audio file with a sampling rate of 44.1 kHz and 16-bit precision is generated. Precise ground-truth labels are also produced during the synthesis process in the form of text files which indicate the location of beats, downbeats and the onsets of each drum type. Fig. 3.1 depicts an example of the audio waveform and the ground-truth labels obtained through the synthesis process.

Apart from rendering the basic rhythmic patterns and combinations of them into synthetic audio files, additional examples were prepared by Luis Jure, which try to simulate real performances. This was carried out taking into account alterations and ornamentations of the *piano* drum pattern and variations of the *clave*

[1] `http://www.lilypond.org/`
[2] `http://www.csounds.com/`

pattern. Besides, the parts of the *repique* drum were constructed from transcriptions of recorded executions by renowned drummers,[3] combining typical motifs and forms. The music scores of two of these synthetic performances are provided in Appendix C.

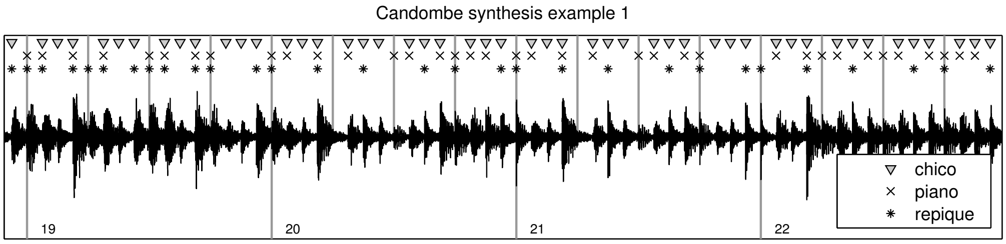


Figure 3.1: Bars 19 to 22 of the synthetic example 1 from Appendix C. Vertical lines in the upper part of the waveform plot indicate the location of beats, whereas the ones in the lower part correspond to the downbeats. Onsets of each type of drum are also depicted with markers.

3.3 Dataset for beat and downbeat tracking

A dataset for beat and downbeat tracking of *candombe* recordings was compiled and annotated for this work. It was released to the research community with the publication of [219], being the first resource of this type available for *candombe*.[4]

The recordings were collected in the context of musicological research over the past two decades, encompassing various recording sessions conducted by Luis Jure in 1992 and 1995, and one recording session carried out as part of this thesis work, which is described in detail in Section 3.4. All the recordings were produced in studio using professional audio equipment. The audio files of the dataset are stereo with a sampling rate of 44.1 kHz and 16-bit precision.

The dataset comprises 35 complete performances by renowned players, totalling over 2 hours of audio, in groups of three to five drums. A total of 26 *tambor* players took part in the recording sessions, belonging to different generations and representing all the important traditional *candombe* styles (the list of performers is provided in Appendix A).

The location of beats and downbeats was annotated by Luis Jure, adding to more than 4700 downbeats. The annotation process involved real-time tapping to the recordings, followed by the adjustment of the placement of some beats manually. The annotations were released as comma-separated value files (.csv) in which data is stored as plain text, following the file format described in Section 3.4.4.

[3] Pedro 'Perico' Gularte, Segio Ortuño, Waldemar 'Cachila' Silva and Wilson Martirena.
[4] Available from `http://www.eumus.edu.uy/candombe/datasets/ISMIR2015/`.

3.4 An audio-visual database of performances

This section describes the creation of an audio-visual database of *candombe* performances in the context of this research. It is intended for computational musicological studies, and includes annotations of metrical information (beat and downbeat), temporal location of strokes and sections. A discussion on the technical requirements and tackled challenges is provided, in the hope that it can be useful for other researchers involved in producing datasets for computational analysis of music.

The goal was to generate useful data that could serve to analyse a percussion performance in an efficient, affordable, and non-invasive manner. Although different approaches were considered such as motion-capture systems and various types of sensors [283], audio and video recordings were favoured as non-intrusive means of capturing the performance. Thus, five renowned *candombe* drummers were recorded on a multi-track audio system and simultaneously filmed by three video cameras. Besides, in order to properly register the fast movements of the percussionists, high frame rate videos were also produced, by means of two additional affordable cameras adapted for this purpose.

The author of this thesis acted as producer of the session and was responsible for the audio recording. In addition, he supervised the group in charge of the high-speed cameras. Luis Jure was the music curator, which involved among other tasks, selecting the group of performers and conducting the recording session.

3.4.1 Group of performers

Several factors had to be taken into account when selecting the group of players to participate in the session. The first criterion was to pick players of the highest level, and representative of the most authentic tradition of *candombe* drumming. But in addition to the individual quality of each player, the balance of the ensemble as such also had to be carefully planned. The first basic condition was that all the drummers had to belong to the same *barrio*, in order to guarantee stylistic compatibility; but musical and personal understanding and affinity among all the performers was also essential. One final consideration was including some performers proficient in more than one type of *tambor*, in order to have a wider range of combinations with the limited number of players that could be convened.

Eventually, a group consisting of five performers was assembled, all of them belonging to families of long-standing tradition in the community of *barrio Palermo* (Ansina): Gustavo Adolfo Oviedo Gradín (b. 1953), Fernando Silva Pintos (b. 1955), Sergio Ariel Ortuño Priario (b. 1966), Héctor Manuel Suárez Silva (b. 1968), and José Luis Giménez García (b. 1969). Besides being widely acknowledged as outstanding players in one or more types of *tambor*, and having led the *cuerda de tambores* of several *comparsas*, they all have vast experience as professional musicians and performing artists in different settings of popular and orchestral music.[5]

[5] Most notably in the "Suite de Ballet según Figari" (1952) by Uruguayan composer Jaurés Lamarque Pons (1917–1982), an orchestral piece that includes Candombe drums in the last movement.

Gustavo Oviedo, one of the most important and influential players of *tambor piano* of the last decades, has been for many years the leader of the drums of *barrio Palermo*. Together with his late brother Edinson "Palo" Oviedo, he has directed two important *comparsas* of the neighbourhood: Concierto Lubolo and Sinfonía de Ansina. Although a strong player of *tambor piano* in a *llamada de tambores*, Fernando "Hurón" Silva is mainly known as a virtuoso player of *repique* in small groups. Together with the Oviedo brothers, he co-led the *comparsa* Concierto Lubolo in the 1980s and early 1990s. At that time, the three of them formed a legendary trio that participated in many important recording sessions of Uruguayan popular music.[6] During the session, Oviedo and Silva only played *piano* and *repique* respectively.

The three remaining performers belong to a younger generation, and the three are known as accomplished players of the three types of drum. Luis Giménez played *chico* and *repique*, while both Sergio Ortuño and Héctor Manuel Suárez (primary *repique* and *piano* respectively), played the three drums in different takes. The number of takes of each drum for the different performers is provided in Table 3.3.

3.4.2 Recording session

In order to encompass both the documentary and research purposes, the recording session had to be carefully prepared. It was devised taking into account a wide range of possible MIR problems, such as drum event detection, audio source separation, automatic transcription, beat-tracking, and audio-visual music analysis.

Several pilot studies were carried out to evaluate technical requirements and to test different solutions. A preliminary recording session—with a group of four professional percussionists—was conducted in studio, which allowed to foresee some difficulties and to identify shortcomings beforehand.

The selected venue was a modern medium-size concert hall, called Zavala Muniz, which is part of the Solís Theatre building, located in Montevideo. Among other facilities, it provided appropriate lighting and acoustic conditions. The recording session took place on September 6th 2014, and lasted for about 6 hours, involving a crew of a dozen people.

Audio recording

The audio set-up was selected to produce two main different outcomes: a stereophonic recording of the ensemble and separate audio channels of each drum. The aim of the former is to provide a realistic spatial sound image of the scene, with good localization of sources plus the effects of the room acoustics [34]. Conversely, the purpose of the separate channels is to record only the direct sounds coming from a given drum, with no interference from the others. This is intended to facilitate the analysis and transcription—either automatic or manual—of the individual performances. Additionally, in this way, the whole set of channels constitutes an appropriate research framework for sound source separation.

[6] Among many others, with noted Uruguayan musician Jaime Roos.

The selected microphone set-up allows for different options to create a stereophonic audio mixture. Two coincident microphone techniques [97] (in which spatial image is only based on intensity differences) were adopted: a pair in mid-side (MS) configuration and a pair in XY pattern. The XY microphone pair was mainly intended as a backup and was connected to a separate portable digital audio recorder. Unlike the fixed spatial image of the XY pattern, the MS technique gives some flexibility to adjust the width of the stereo spread after the recording is finished. Besides, the mid MS channel grants monoaural compatibility, whereas collapsing the XY tracks into mono can result in some phase cancellation in high frequencies. Finally, since both pairs were placed close to each other, this audio redundancy can be used to study the influence of the microphones on problems such as source separation or sound recognition.

In addition, a spaced microphone technique [97] (which also captures time-of-arrival differences) was applied, using a pair of omnidirectional microphones in AB configuration. The AB pair was placed in a T-shape array with respect to the coincident stereo pairs, so that they form a three-point pick-up pattern called *Decca tree* [97] (after the record company). By combining the coincident and spaced microphone techniques into a stereo downmix, a good compromise between spatial sense, stereophonic image and centre definition can be obtained. Furthermore, the low frequency response is improved by the contribution of the omnidirectional microphones, compared to that of the directional microphones used in the coincident pairs.

Separate audio channels were obtained by close-miking each drum and recording them to independent tracks. Yet, achieving good source separation given the high sound pressure level produced by the drums turned out to be challenging. Previous tests were conducted with acoustic panels standing in-between musicians (called gobos) to reduce spill from one instrument into the other spot microphones. Although effective, they interfere with a natural performance and were discarded. Eventually, a set of dynamic microphones—tailored for percussion instruments—yielded the best results among the tested options. They are able to handle high pressure levels and their moderate sensitivity prevents them from picking too much sound from other drums.

Table 3.1, summarizes the microphone set-up and gives additional details, including manufacturer and model of microphones and recorders. All the mics chosen for recording the spatial sound image were of condenser transducer type, for their higher sensitivity to capture distant sounds, reverberation and nuances. Audio was recorded at a sampling rate of 48 kHz and 24-bit precision. An outlook of the stage is shown in Fig. 3.2, during the performance of a five-drum ensemble.

Video recording

For documentary purposes a conventional video filming set was used. It comprises three video cameras equipped with professional grade lenses, two Canon 7D and a Sony alpha-99. Video was recorded at 24 frames per second (fps) and 1920x1080 pixel resolution (HD, H.264). One of the cameras always captured a wide shot of the ensemble, while the other two focused on closed-up views of the *repique*

Sound Devices 788T-SSD				
channel 1	Schoeps	MK 4	cardioid	mid
channel 2	Schoeps	MK 8	figure 8	side
channel 3	Sennheiser	E604	cardioid	spot
channel 4*	Sennheiser	E604	cardioid	spot
channel 5	Sennheiser	E604	cardioid	spot
channel 6	Sennheiser	E604	cardioid	spot
channel 7	AKG	C414	omni	left
channel 8	AKG	C414	omni	right
Tascam HD-P2				
channel 1	Neumann	KM 184	cardioid	left
channel 2	Neumann	KM 184	cardioid	right

* Neumann TLM 103 in isolated strokes and solo performances.

Table 3.1: Input list and microphone set-up for audio recording.

Figure 3.2: Stage outlook during a five-drum performance.

and *piano* players, as depicted in Fig. 3.3. A traditional clapperboard was used for embedding a visual and aural reference to synchronize the different audio and video sources.

Due to the very fast movements of an accomplished percussionist playing *candombe*, the computational analysis of a performance based on video records at conventional frame rates (e.g. 24 fps) is quite limited. For this reason, high-speed cameras were also employed. From a survey of commercially available cameras of this type, the GoPro HERO3+ Black Edition was chosen as an affordable option. Coincidently, it has been recently reported to be used for similar purposes [289]. It is a rugged and compact action camera, primary associated with outdoor sports,

Figure 3.3: Video framing examples, a wide shot of the ensemble (above) and close-up view of the *repique* performer (below).

often attached to helmets or surfboards. Thus, various technical issues had to be tackled to adapt it for the current application.

Based on previous tests, a frame rate of 240 fps and 848x480 pixel resolution was selected. This produced reasonably smooth data and low blur of moving objects, and proved to be suitable for automatic processing. At such high frame rates, fluctuations in the lighting conditions are critical and the power-line flicker effect can arise. Experiments conducted in advance in the concert hall suggested that the change in brightness of the existing incandescent lamps was not troublesome. However, moderate lighting fluctuations—almost not perceivable to the naked eye—are present in the recorded videos, and can hinder the performance of video processing algorithms if not properly dealt with.

Two cameras of the same type were used simultaneously, which provided—apart from hardware backup—a stereo recording that can be used for 3D analysis and scene reconstruction (see Fig. 3.4). This entailed a calibration process with a chequerboard pattern during the recording session.

By rotating 90°the wide angle field of view, a more adequate framing was obtained. The camera has no view finder, so it had to be connected to a display screen in order to check the correct framing of each video take. A remote control allows for basic operation of the device, but the wireless communication has to be enabled, which is a power demanding feature. For this reason, battery endurance was not enough and connection to a power supply became mandatory. Consequently, the need of wiring the camera, for monitoring and powering, prevented the use of the standard housing and mounting accessories. Besides, only relying on the remote control turned out to be not sufficiently robust and access to the buttons had to be granted. Access to the SD card slot was also necessary (which is not possible with the housing), to readily transfer the large files generated at high frame rate. Given all these requirements, a custom camera mounting had to be made, capable of holding two cameras in upwards position, with an adjustable distance between them and providing access to all the necessary connections and buttons. This mounting can be seen in Fig. 3.5, attached to a standard tripod. Because of the fixed focal length of the camera, the tripod had to be placed not very far from the subject to get a detailed view. As a result, only one musician was effectively recorded at high-speed in ensemble performances.

The scene was slightly prepared for research purposes, to aid the application of automatic video processing techniques and to simplify the evaluation of algorithms. This was done in a non-invasive manner, taking care not to alter the sound of the drums or disturb the performer. As shown in Fig. 3.4, the stick and the contour of the drumhead were painted to ease their automatic segmentation. In addition, some fiducial paper markers were attached to the floor, the drum and the performer's body for evaluation purposes. All the drums used were 3D scanned using a Kinect system to accurately register shape data in case it could be useful for future research.

3.4.3 Dataset

During the recording session, three different types of content were registered, namely, isolated strokes, solo performances and drum ensembles. The dataset comprises 51 takes, totalling a duration of nearly an hour and a half. This section provides a short description of how each type of content was rendered, as well as their intended purposes.

Some musicians were asked to change role during the session and played different drums. Besides, for the sake of data variability, different instruments of the same drum type were used. To do that, two instruments of each type (i.e *chico*, *repique* and *piano*) were utilized. This set of drums was previously prepared, by painting the drumhead contour and attaching some fiducial paper markers to the shell and drumhead (as shown in Fig. 3.4). But in addition, the performers were

Figure 3.4: Matching frames of the stereo recording pair.

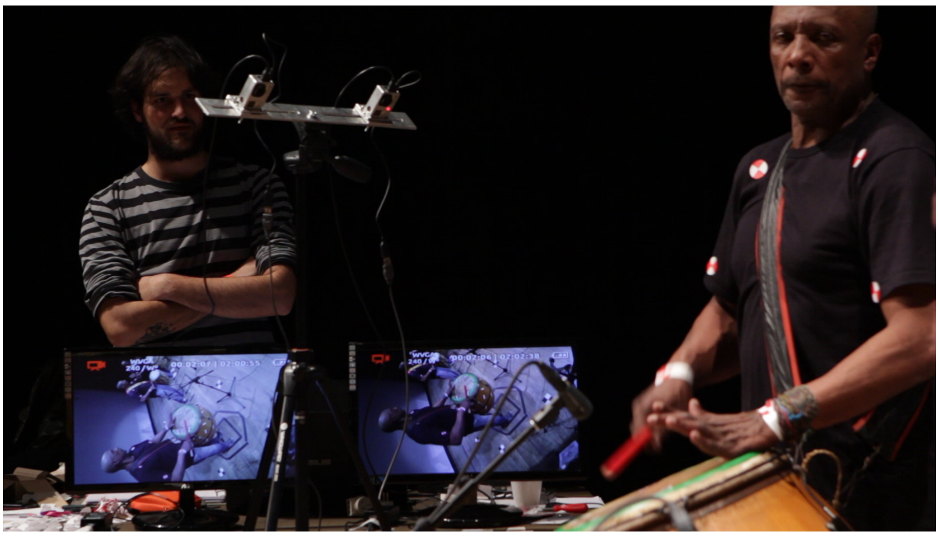


Figure 3.5: High-speed cameras in stereo pair with custom-made mounting, wired to two screens and power supply.

asked to bring their own instruments—which where obviously not prepared—and part of them were involved in recording isolated strokes and drum ensembles. Table 3.2 shows the number of takes of each content type, and Table 3.3 provides the number of takes in which each performer is involved.

Strokes		**Solo**		**Ensembles**	
chico	4	chico	8	three drums	9
repique	4	repique	10	four drums	3
piano	3	piano	8	five drums	2

Table 3.2: Number of takes of each content type.

	Strokes			**Solo**			**Ensembles**		
	chico	repique	piano	chico	repique	piano	chico	repique	piano
Silva		1			3			8	
Suárez	1	1	1	3	2	3	2	1	6
Ortuño	2	1	1	2	2	2	2	8	1
Oviedo			1			3			9
Giménez	1	1		3	3		10	2	

Table 3.3: Number of takes by each performer.

Isolated drum strokes

The musicians were asked to produce the sound of individual strokes separately, and were recorded in turns playing different drums. These strokes are supposed to be the same they would use in a real performance. Therefore, they were requested to render some particular stroke types, but also to include those that belong to their personal repertoire.

The set of isolated drum sounds is intended for two main different uses: firstly, to provide a database of strokes suitable to train and evaluate sound recognition algorithms, which are typically part of automatic transcription systems; secondly, to build a set of audio samples of each stroke type to be used in the sample-based audio synthesis process described in Section 3.2.

Solo drum performances

Although *candombe* is—above all—a collective form of music, part of the session was devoted to recording each musician alone executing the rhythmic patterns of a certain drum. They were asked to play at different tempos and in different styles, and freely developed their improvised parts.

The aim of these recordings is to avoid any interference from other drums, something which is not completely fulfilled by the separate channels of an ensemble recording. In this way, automatic tasks such as onset detection can be carried out without the need to deal with spurious events from other drums. In addition, this allows, for instance, to study spectral timbre features for sound recognition in a more realistic situation compared to that of the isolated strokes. Ultimately, these performances can be contrasted with that of ensembles, in order to investigate to what extent musical behaviours are alike.

Drum ensembles

Ensembles were recorded in groups of three, four and five performers. The first case corresponds to a *candombe* ensemble in its minimal form, that is, one of each of the three drums. In the case of four performers, an extra *repique* was added to the ensemble. Finally, the groups of five drums consist of two *piano*, two *repiques* and one *chico*.

As a result, 14 complete ensemble recordings were produced, that last from about 2 to 4 minutes each, for a total of 40 minutes. The different groups, which involved several combinations of the same musicians (as well as some of them in different roles), yielded various sorts of interactions and performance types. For instance, different characteristics typologies were executed for beginning the performance, including *piano* anacrusis and several variations of the *clave* pattern. In the same way, different dynamics and tempo—as well as variations of them—were also performed.

Care was taken to alternate the type of drum in front of the high-speed cameras, in order to have a balanced record of *repique* and *piano* performances in ensemble. The resulting separate audio channels are adequate for several automatic tasks such as onset detection, transcription and source separation. The amount of interference from other drums depends on the distance between performers. For this reason, the best results were obtained for the recordings of three drums, in which the performers were farther apart.

Evidently, the most important aspect provided by ensemble performances compared to the previous recordings is the interplay between musicians. This encompasses call-and-response interactions, alternations of musical roles, variations in dynamics, temporal synchronization, and collective modulations of tempo, just to name a few. Among the various aspects that can be studied from these recordings, the different forms of interaction and the entrainment processes involved are some of the most appealing. Besides, the different sorts of musical embodiment are also of great interest. For example, foot tapping can be observed in the wide shot of the ensemble (see Fig. 3.3-above).

3.4.4 Annotations

Manual labelling of music is notably a laborious process, for which software tools have been developed [66] and methodologies and file formats have been proposed [182]. The annotations are very useful for both musicological studies and the development of music information technologies. This section describes some of the efforts that were accomplished—and others that are still being conducted—regarding the annotation of this dataset.

Beat and downbeat annotation

The 14 ensemble takes were annotated with beat and downbeat labels by Luis Jure. These recordings and labels are part of the dataset used for beat and downbeat tracking, which is described in Section 3.3. Besides, the annotation of solo performances is currently being carried out and will also be released. It has already been completed for the *piano* solo recordings, which were used within this work in rhythmic pattern analysis and downbeat tracking.

The annotations are provided as comma-separated values files (.csv) in which data is stored as plain text, as shown in Fig. 3.6. The files contain two columns, and each line corresponds to a beat. The values in the first column are the time instants of beats in seconds. The numbers on the second column are two values separated by a dot, which indicate the bar number and the beat number within the bar, respectively. For instance, 1.1, 1.2, 1.3 and 1.4 denote the four consecutive beats of the first bar. Hence, each label ending with .1 corresponds to a downbeat.

```
1.78875,   1.1
2.31252,   1.2
2.83536,   1.3
3.35711,   1.4
3.87728,   2.1
4.39545,   2.2
4.91217,   2.3
5.42815,   2.4
...    ,   ...
```

Figure 3.6: Example of the format of the beat annotation files.

Annotation of onset and stroke type

Another effort was undertaken to annotate the location of each onset and its stroke type. To this effect, solo performances and separate channels of the ensemble recordings are used. The annotation process is facilitated by automatically locating events through a standard onset detection algorithm based on spectral flux [94], which is described in Section 4.3. The resulting events are manually validated and/or corrected, by inspecting the audio and video files. Finally, a certain class label representing the stroke type is manually assigned to each onset. These annotations can be used in research on onset detection, sound recognition and automatic transcription.

The author of this thesis extracted, manually checked, and corrected when necessary, the location of all the onsets from the separate channels of the 14 ensemble recordings, totalling about 40000 events. This was used in the study of micro–rhythmical properties of *candombe* drumming, that was reported in [159] and [251], and is described in Section 5.3. In addition, the author annotated eight recordings of *repique* solo performances from the preliminary studio session, indicating onset location and stroke type out of six different classes, for a total of more than 4000 strokes. The annotations, audio files and high-speed videos, were used for the development of an audio-visual transcription system which was reported in [197]. Eventually, all this material will be available to the research community.

Miscellaneous annotations

Several other sorts of annotations were produced—some are still being generated—that depend on the type of research problem addressed and the particular characteristics of the dataset. For example, an experiment was conducted attempting to identify those temporal segments of a *repique* performance when the *clave* pattern is played, which was reported in [252]. This involved manual annotation of separate tracks from ensemble recordings and yielded very informative data about the interaction of two *repique* performers playing together, as described in Section 4.4.4. The annotated sections in which the *clave* pattern is played were used for the study of their different types and variants, as reported in [160] and detailed in Section 5.2.

Regarding to the high-speed videos, precise annotation of the most important objects that appear in the scene—namely, the drumhead, the stick and the performer's hand—is necessary for the development and assessment of automatic detection algorithms [197]. This is a very time consuming task, and efficient labelling mechanisms are being developed based on the fiducial markers.

3.5 Conclusions

This chapter introduced the datasets used along this thesis work and—since they were not previously available—described the necessary efforts fulfilled to generate them. This involved an important amount of work, including collecting and labelling audio files, conducting various recording sessions and developing custom software tools for audio synthesis. A detailed description of the tasks, requirements and technical challenges involved in producing the audio-visual database of *candombe* performances was provided.

Apart from the documentary value and its potential in educational activities, there are different research strands that can benefit from the produced datasets. A continuous effort to extend the existing annotations of the dataset has to be undertaken, in order to expand its possible applications. As a result of this research, a database of audio recordings and annotations for beat and downbeat tracking was released. As part of the future work perspectives, the whole data produced—including the annotations—will be ultimately available to the research community to foster reproducible research.

Chapter 4

Audio features

This chapter reproduces some passages of articles [252] and [254], and also includes all necessary modifications and additions in order to put the work in the context of this thesis.

4.1 Introduction

The extraction of musically meaningful content information via automatic analysis of audio recordings has become an important research field in audio signal processing. Most of this research has concentrated on pitched instruments, and only in the past decade percussion instruments have gained interest, mainly focusing on the standard pop/rock drum kit [113]. The striking of a drum membrane produces a very short waveform that can be modelled as an impulsive function with broadband spectrum [114], whose accurate characterization and analysis is a challenging problem in signal processing. First, some method has to be applied to automatically find the occurrence of sound events (onsets) and to precisely determine its temporal location. This is usually implemented in the form of a detection function that emphasizes the onset of notes by detecting changes in some properties of the audio signal, such as the energy content in different frequency bands, namely the spectral flux [94]. Besides, features computed from the audio signal can provide additional information about an event, such as the type of percussion instrument which produced it, or the particular class of stroke articulated. Both tasks—onset detection and sound classification—are of paramount importance in several applications, from computer-aided musicology to automatic music transcription.

Many techniques for onset detection have been proposed, a good deal of which are reviewed in [37, 76, 94]. The most typical methods are simple to implement and have low computational complexity, usually involving spectral and/or phase information alone [50]. Current state-of-the-art methods for onset detection are based on a probabilistic model and incorporate a recurrent neural network with the spectral magnitude and its first derivatives as input features [100]. However, for percussive onsets their performance is similar to that of traditional methods, such as those based on the spectral flux, which are less computationally demanding and do not require a training phase as in the case of neural network methods. Detection of percussive onsets is claimed to be a solved problem [53], given that, for instance, state-of-the-art methods achieve F-measure values greater than 0.90 on the 30 solo drum excerpts of the MIREX evaluation campaign.[1] Yet, there is still room for improvement towards even more robust and versatile tools.

Regarding sound classification, even if the problem of dealing with isolated sound events is widely studied [140], the performance of the available methods largely decreases when simultaneous sounds and real performances are considered [138, 272]. Existing approaches for percussion transcription can be roughly divided into two types [113]. Most of the proposed solutions apply a pattern recognition approach to sound events. Firstly the audio signal is segmented into meaningful events, either by detecting onsets or by building a pulse grid. Then, audio features are computed for each segment, usually to describe spectral content and its temporal evolution [113, 138, 140, 230]. Finally, the segments are classified using pattern recognition methods [230, 279]. The other usual approach is based on segregating the audio input into streams which supposedly contain events from a single percussion sound class, by means of signal separation techniques [67, 122], or simply by sub-band filtering [208]. After that, a class is assigned and an onset detection procedure is applied to each stream. In fact, there are even some approaches that do not fall into any of the previous categories, either because they combine aspects of both [231], or because they aim at the detection of high-level rhythm patterns [282]. Other distinctions can be made, such as whether the classification is supervised or not, and whether it takes or not high-level musicological information into account [113].

In this chapter, the audio features used throughout the thesis are presented. A typical approach is adopted based on the spectral flux, which is well-suited for dealing with the percussive events at hand. In Section 4.2, the audio feature extraction process and some representations built upon it are described. Then, the usefulness of the features is assessed in the context of onset detection and classification of *candombe* drum sound events, in Sections 4.3 and 4.4 respectively. An approach that models for sound classification the same type of spectral flux features employed in onset detection is tested for recognizing drum sounds in audio signals. Two type of experiments are reported involving recordings of real performances, one which aims at finding the predominant *candombe* drum heard in an audio file, while the other attempts to identify those temporal segments within a performance when a given sound pattern is played. The experiments address audio files in which a predominant instrument suffers the interference from some others. This type of audio file could be either the result of a signal separation technique as previously described, or coming from a microphone placed close to an instrument when a multi-instrument performance is recorded. The latter situation is common practice in some music productions or musicological field studies [239], and is the case of the dataset considered in the reported experiments, which was introduced in Section 3.4. Finally, the chapter ends with some discussion on the audio features adopted and on the results of the reported experiments.

[1] `http://www.music-ir.org/mirex`

4.2 Audio feature extraction

The audio features applied throughout this research work are based on seizing the changes in the spectral magnitude of the audio signal along different frequency bands. This typical approach can be tracked back to pioneering work such as [199], and is usually referred to as spectral difference or spectral flux [94]. In the following, the classical formulation is introduced and some variants are described, particularly frequency scaling and log-magnitude whitening, along with a discussion of its potential impact on the percussive sound events of interest. Then, some processing steps used for the analysis of rhythmic patterns, namely normalization, time quantization and sub-band analysis, are described. Finally, a representation in the form of a feature map is proposed as a means to study similarities and differences among the rhythmic patterns as well as their evolution over time.

4.2.1 Spectral flux

The first step to compute the spectral flux of a discrete-time audio signal $x[n]$ is calculating its Short-Time Fourier Transform (STFT), which is defined as

$$X(m,k) = \frac{1}{N} \sum_{n=0}^{N-1} w[n-mh]\; x[n]\; e^{-j\frac{2\pi}{N}kn}, \tag{4.1}$$

where m stands for the index of the signal frame being analysed, k is the frequency bin index, h is a hop size in samples and $w[n]$ is a smoothing window, for instance a Hann window, such that $w[n] = 0 \;\forall\, n$ outside the interval $0 \leq n < N$. Then, the magnitude of the short-time spectra (i.e. the spectrogram) is time-differentiated via first-order difference and the resulting sequences are half-wave rectified to consider only positive magnitude changes. Finally, the detection function is obtained by summing along all frequency bins. This can be expressed as

$$SF(n) = \sum_{k=0}^{N/2} H\left(|X(n,k)| - |X(n-1,k)|\right), \tag{4.2}$$

with $H(x)$ denoting the half-wave rectification function, i.e. $H(x) = \frac{x+|x|}{2}$.

In principle, the feature value is high when a stroke has been articulated and close to zero otherwise. But in addition, the detection function also carries some information on the type of articulation. For instance, an accented stroke produces a higher feature value compared to a muffled one, since the spectral change is more abrupt and typically encompasses a larger frequency bandwidth.

Logarithmic magnitude

A frequent preprocessing step in the computation of the spectral flux is compressing the magnitude of the STFT before differentiation by using a logarithmic function. This better correlates with human perception of loudness, and was reported to be useful for onset detection [167]. Note that taking the derivative of the logarithmic magnitude of the STFT is equivalent to dividing the derivative by the magnitude of the STFT, which corresponds to a relative differentiation. This is consistent with psychoacoustic knowledge, since the perceived increase produced by a certain change in the amplitude of the signal depends on its initial level: the same amount is more noticeable when the signal is soft than when it is loud [167]. Besides, whitening of the spectrum can improve onset detection [277], and the use of a logarithmic function can be regarded as a kind of whitening which emphasises higher frequency bands.

The amount of compression can be controlled by a parameter λ (a way to customize the norm applied), and the addition of a constant value of 1 avoids numerical issues, all which leads to the following expression:

$$X_{\log}(n,k) = \log\left(\lambda\,|X(n,k)| + 1\right). \tag{4.3}$$

Frequency scaling

A frequency scaling of the spectrogram by a filter bank is a typical variation of the previously described process [50]. The linearly spaced frequency bins of the STFT are combined into fewer bands whose centre frequencies follow a scale that better approximates human auditory resolution, i.e. coarser resolution at high frequencies and finer resolution at low frequencies. Among the most frequently used frequency mappings are the Mel and the Bark scales [101]. Most of the rhythmic information of a music audio signal is preserved in spite of the dimensionality reduction produced by the filter bank processing [263], which has been shown to be advantageous for onset detection [76, 167]. Usually a total number of bands B of about 20 to 80 is used [50, 52].

Considering a filter-bank $FB(k,b)$, where k is the bin index of the linearly spaced frequency scale of the STFT and b is the number of the filter, the filtered spectrogram can be expressed using the dot product as

$$X^{\text{filt}}(n,b) = |X(n,k)| \cdot FB(k,b). \tag{4.4}$$

This work adopts a bank of overlapping triangular filters as depicted in Fig. 4.1, whose centre frequencies are equidistant in the Mel scale. The filters are normalized so that no emphasis on higher frequencies is produced by their different bandwidths. The Mel scale mapping can be computed according to,[2]

$$f_{\text{Mel}} = 2595 \log_{10}\left[\frac{f_{\text{Hz}}}{700} + 1\right]. \tag{4.5}$$

An example of the steps involved in the computation of the spectral flux is provided in Fig. 4.2 for a short excerpt of a *repique* performance recorded in studio. The influences of the logarithmic magnitude whitening and the Mel scale mapping can be appreciated by comparing the three different feature functions depicted. Clearly, the third strategy ($SF^{\text{filt}}_{\log}$, i.e. logarithmic magnitude and frequency scaling) produces better defined peaks without increasing the noise floor.

[2] The actual implementation matches Slaney's Auditory Toolbox mapping [273].

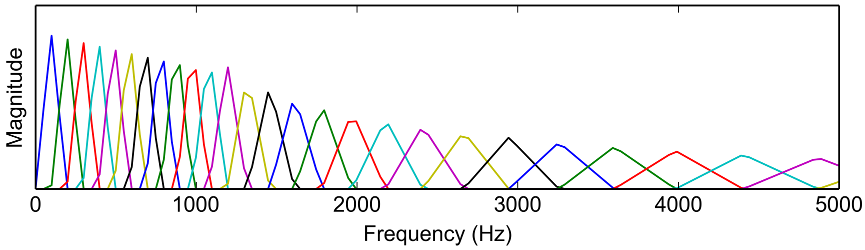


Figure 4.1: Mel scale filter bank used for the computation of features.

Normalization

Apart from indicating the occurrence of sound events and carrying some information about the type of articulation, the spectral flux also captures the long-term dynamics of the audio signal, since louder portions take higher values compared to quieter ones. Therefore, different types of normalization are usually applied to the feature function. For certain tasks, such as onset detection, a global normalization—implemented by simply dividing the detection function by its maximum value—proved to be appropriate for our purposes, as long as an adaptive threshold is used (see Section 4.3).

However, for the analysis of rhythmic patterns and for addressing the beat and downbeat tracking problem—as described in Chapters 5 and 6 respectively—the feature must preserve the local intensity variations that characterize the different articulations within a pattern, but at the same time must not be influenced by the long-term fluctuations in dynamics that may arise throughout a performance.

For this reason, the detection function was normalized by the p-norm within a local window centred at the current time frame as

$$\overline{SF}(n) = \frac{SF(n)}{\sqrt[p]{\sum_{m=-\Delta}^{\Delta} |SF(n+m)|^p}}, \tag{4.6}$$

where p controls the type of norm applied and Δ determines the window length. The former has to take a high value (for instance, $p = 8$ was used for most of the reported results), so that if the feature in the current frame is close to the highest value within the window it is normalized to 1. For the normalization to behave as desired, the Δ parameter must be selected such that several sound events lay within the window. This can be implemented by considering Δ to be proportional to the *tatum* period of the performance, that is

$$\Delta = T\,\tau, \tag{4.7}$$

where τ stands for the *tatum* period in samples and T takes an integer value such that $T > 0$. In most cases, a value of $T = 4$ was used, which corresponds to a window of half a bar centred at the current time frame. Note that the *tatum* period is estimated either from the labelled *tactus* pulses for the analysis of rhythmic patterns (see Chapter 5), or directly from the audio signal when beat and downbeat tracking is tackled (as described in Section 6.2.2).

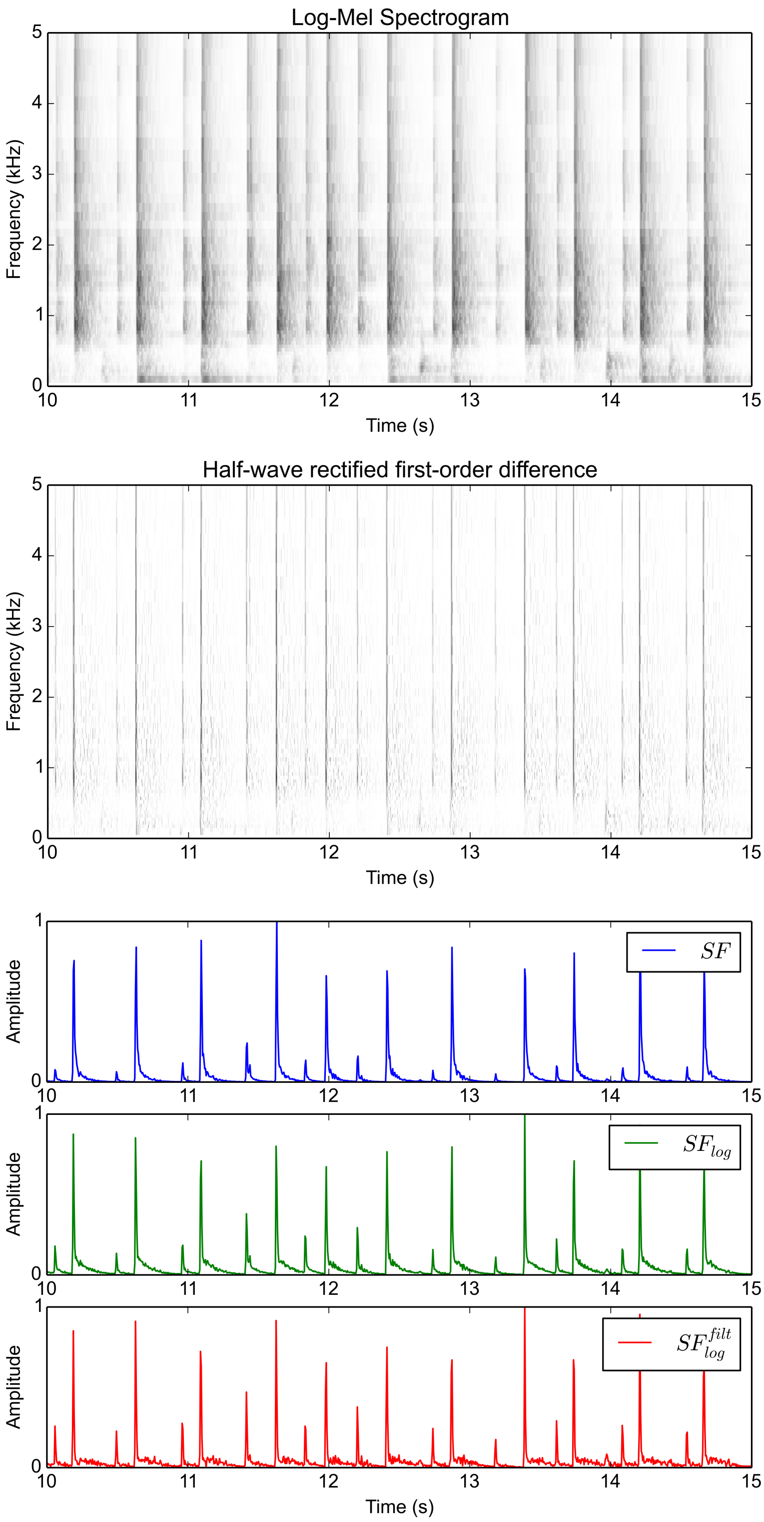


Figure 4.2: Example of spectral flux computation for a short excerpt of a *repique* performance, and comparison of the influence of the logarithmic magnitude and the frequency scaling steps.

Time quantization

For the analysis of rhythmic patterns the feature signal is time-quantized by considering a grid of *tatum* pulses equally distributed within the labelled *tactus* beats. The corresponding feature value is taken as the maximum value of the feature signal within a 100-ms window centred at the frame closest to the *tatum* instant.[3] This yields 16-dimension feature vectors in which each coordinate corresponds to a given *tatum* pulse within the bar. Although intended to illustrate the sub-band analysis of the next section, Fig. 4.3 provides an example of time quantization of the feature function, in which the selected values are indicated by crosses.

4.2.2 Sub-band analysis

Given the distinct registers of the different drum types and the high frequency content of the *madera* sound of the *clave* pattern, a rough separation of the rhythmic patterns was pursued by sub-band filtering as in [208]. This was implemented by summing the spectral flux along different frequency bands, as in [174].

Some controlled experiments were conducted in order to test the validity of this separation approach and to determine the frequency-band boundaries. To that end, synthetic test audio signals rendering different rhythmic patterns were produced with the software tools described in Section 3.2. Fig. 4.3 shows the spectral flux in three different frequency bands for two bars of a synthetic signal comprising *piano*, *chico* and *clave* patterns (as in Fig. 2.13). The articulated beats of each pattern are depicted with dots for the low, medium and high frequency bands, respectively. It can be seen that peaks in the feature signal approximately match the synthesized patterns. Also, if the median of the feature values for each *tatum* beat within the bar is computed along the whole audio file, the resulting feature profiles are consistent with the synthesized patterns, as shown in Fig. 4.4a.

To take into account a more realistic scenario, the same type of analysis is presented in Fig. 4.4b for a 30-second excerpt from a field recording of a *candombe* performance. The ensemble is composed of one drum of each type, and the *repique* plays a *clave* pattern in between improvised phrases. It can be noticed that the prototypical patterns of *piano*, *chico* and *clave*, which are depicted with dots as reference, show some differences to the feature values. However, careful inspection of the audio file reveals that the feature profiles in fact correspond to the actual patterns played in the recording. In particular, the *chico* pattern is played by also softly articulating the first beat and with a very accentuated hand stroke at the second, whereas a variation of the *clave* pattern is performed in which the third stroke is shifted to the next *tatum* beat (as in a rumba *clave*). Considering all the experiments conducted, even thought interference between the different bands may arise in some cases, the separation approach proved to be quite effective and thus motivated its application to the analysis of rhythmic patterns and to the problem of beat and downbeat tracking.

[3] At a high tempo value of 140 bpm, this quantization window approximately corresponds to half the duration of the tatum period to each side of the tatum beat location.

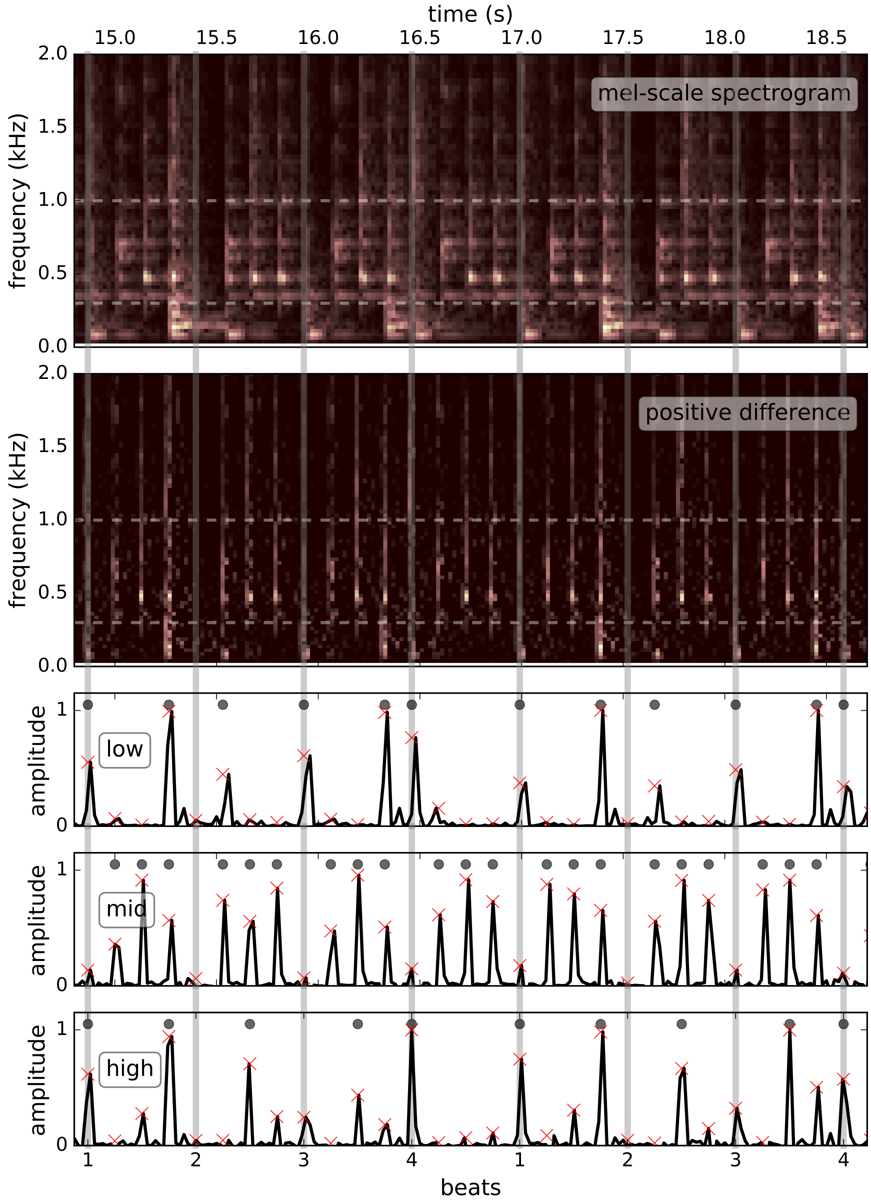
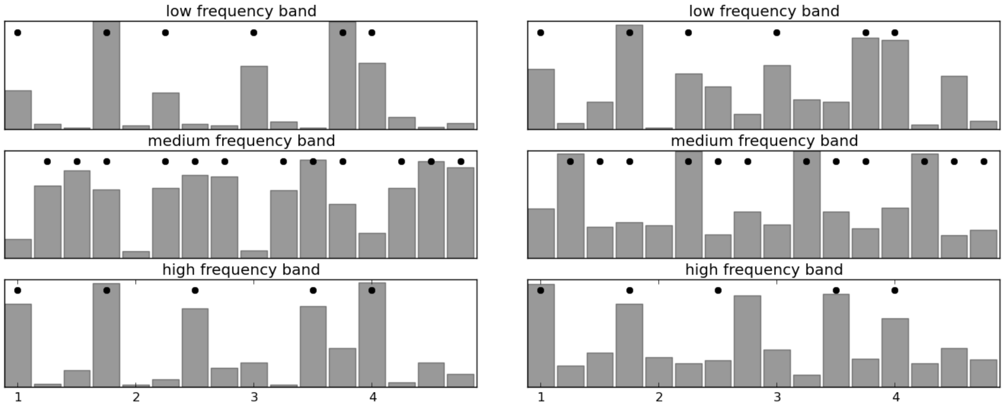


Figure 4.3: Example of audio feature extraction: magnitude of the Mel-scaled short-time spectra (top), half-wave rectified first-order difference (middle), and accentuation feature in three different frequency bands (bottom). The frequency bands limits are, low: up to 200 Hz, medium: 400 to 1000 Hz, and high: 1000 to 1600 Hz. The audio signal comprises synthetic *piano*, *chico*, and *clave* patterns. The articulated events of each pattern are depicted with dots in the accentuation feature plots for the low, medium, and high frequency bands, respectively. Frequency band limits are shown with dashed lines and beat locations are depicted with vertical lines. It can be seen that peaks in the feature signal approximately match the corresponding patterns. The time-quantized feature values are indicated with crosses.



(a) synthetic signal (b) field recording

Figure 4.4: Feature profiles for a synthetic audio file (a) and a field recording (b), obtained as the median of the feature values for each *tatum* beat within the bar.

4.2.3 Map of feature patterns

A representation in the form of a map of bar-length rhythmic patterns is proposed in this work, which is straightforwardly obtained by building a matrix whose columns are consecutive feature vectors. An example of this type of map, computed from a complete performance, is provided in Fig. 4.5. The horizontal axis corresponds to the bar index, while the vertical axis is the *tatum* beat, increasing upwards as convention. The features were computed using only the low-frequency band and then warped into bar-length patterns using the beat and downbeat manual labelling. Therefore, the columns of the map virtually correspond to each of the bar-length patterns performed by the *piano* drum along the whole recording.

This representation enables the inspection of the patterns evolution over time, as well as their similarities and differences, in a very informative way. Note that if a certain *tatum* pulse is articulated for several consecutive bars, it will be shown as a dark horizontal line in the map. Conversely, changes in repetitive patterns are readily distinguishable as variations in the distribution of feature values. The example shown in Fig. 4.5 is analysed in detail in Section 5.2.2.

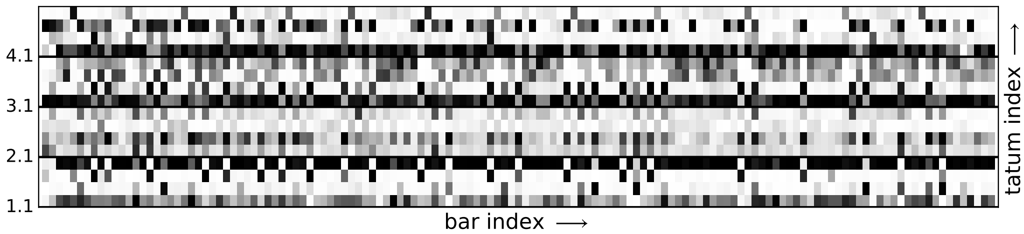


Figure 4.5: Map of bar-length patterns for a recording of the dataset of Section 3.3. Vertical axis ticks indicate *tatum* pulses and horizontal axis ticks correspond to bar index numbers.

4.3 Onset detection

In this section the usefulness of the spectral flux features previously introduced is assessed for the onset detection task considering *candombe* drum sounds. The onset detection methods are usually divided into three steps: signal pre-processing, computation of the onset detection function, and peak picking [100]. The detection function employed in this case is based on the spectral flux and has been already introduced, as well as the logarithmic magnitude and Mel-scale mapping pre-processing steps. The peak selection method has a strong influence on the results obtained [258], and the one adopted herein is discussed in the following.

4.3.1 Peak picking

Following a method proposed in [94] and later modified in [50], a set of simple peak selection rules were implemented in which onset candidates, apart from being a local maximum, have to exceed a threshold that is a combination of a fixed and an adaptive value. Thus, the spectral flux has to fulfil the following two conditions:

$$SF(n) = \max\{SF(n-\hat{\omega}_{\text{pre}} : n+\hat{\omega}_{\text{pos}})\} \tag{4.8}$$

$$SF(n) \geq \text{mean}\{SF(n-\bar{\omega}_{\text{pre}} : n+\bar{\omega}_{\text{pos}})\} + \delta \tag{4.9}$$

where δ is a fixed threshold and the ω parameters determine the width of the moving average and moving maximum filters, i.e. the number of previous and subsequent points involved. Note that if the ω_{pos} parameters are not zero the peak selection is not causal. A global normalization was added by dividing the detection function by its maximum value, which simplifies the selection of the fixed threshold value across different performances. An example of the peak selection method is provided in Fig. 4.6 for a *repique* studio recording, which depicts the adaptive threshold and the moving maximum condition.

An additional restriction is sometimes applied to limit the minimum time span between two consecutive onsets [50]. However, the local maximum condition can be tuned such that among several events occurring in a short time interval only the most prominent is reported. In a *candombe* drumming performance there are several situations of this kind. For instance, the *flam* is a stroke which consists of two single strokes played almost together by alternating the hand and stick, and is commonly played by *piano* and *repique* drummers. There is also an ubiquitous *repique* stroke in which the stick hits the drumhead several times in a short time interval, namely a *bounce*. The moving maximum condition proved to be effective to deal with these situations.

4.3.2 Dataset

For testing the onset detection method a dataset comprising eight solo performances of the *repique* drum was used. They were recorded in the preliminary studio session by four professional percussionists (see Section 3.4.2), with a sampling rate of 44.1 kHz and 16-bit precision. The annotation of these recordings

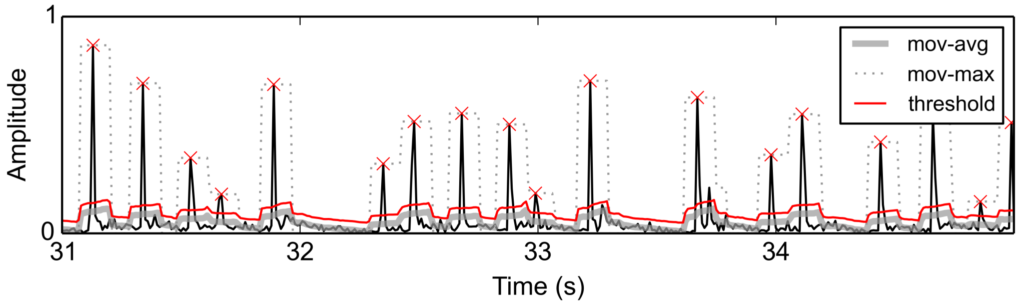


Figure 4.6: Example of the peak detection method. Selected onsets are marked with crosses.

	Fmeasure	precision	recall
SF	0.979	0.980	0.977
SF_{log}	0.987	0.985	0.988
$SF_{\text{log}}^{\text{filt}}$	0.987	0.986	0.987

Table 4.1: Onset detection results for the different detection functions.

was carried out by the author and involved using audio and video, for a total of 4132 onsets. Each onset was labelled to a certain stroke type out of six different classes. This dataset was also used in the evaluation of a multimodal approach for percussion music transcription from audio and video, which was reported in [197].

4.3.3 Experiments and results

The audio signals were processed in frames of 20 ms length, using a Hann window and a hop size of 10 ms. The number of Mel frequency bands was set to $B = 80$ and the compression parameter to $\lambda = 10^4$. The ω parameters were manually set in order to avoid false positives for strokes that involve several events in a short time interval, a wide range of values yielding very similar results. For the reported experiments their values are $\hat{\omega}_{\text{pre}} = \bar{\omega}_{\text{pre}} = 50$ ms and $\hat{\omega}_{\text{pos}} = \bar{\omega}_{\text{pos}} = 70$ ms.

An onset was considered correct if there was an anotation within a window of ± 25 ms centred at the detected event. In this way, the false positive fp, false negative fn, and true positive tp rates were calculated for the whole dataset. Then, precision, recall and Fmeasure were computed as

$$\text{precision} = \frac{tp}{tp + fp}, \quad \text{recall} = \frac{tp}{tp + fn}, \quad \text{Fmeasure} = 2\,\frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}. \qquad (4.10)$$

The fixed threshold δ was varied from 0 to 1 producing the ROC curves [102] depicted in Fig. 4.7, for the spectral flux computed without any pre-processing SF, and with the addition of the logarithmic magnitude SF_{log} and the Mel-scale mapping $SF_{\text{log}}^{\text{filt}}$. The best performing configuration for each case was selected considering the maximum of the Fmeasure, and is reported in Table 4.1.

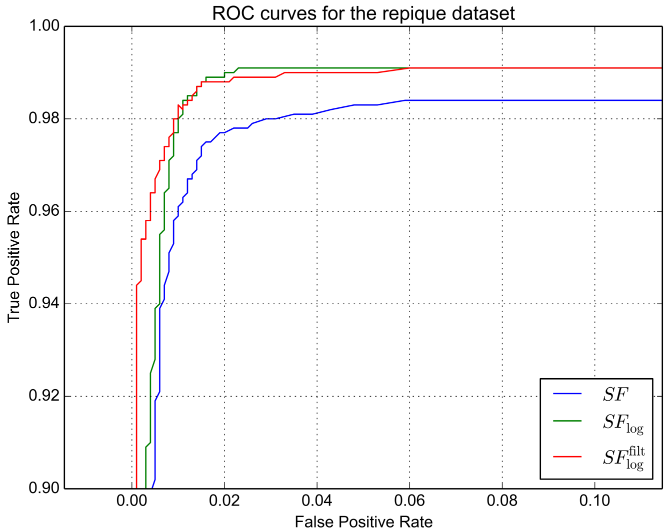


Figure 4.7: ROC curves comparing the different detection functions.

The performance of the onset detection is notably high. This is probably favoured by the fact that only solo performances were considered. In the case of ensemble performances recorded into separate audio channels, there is some interference from one instrument into the other spot microphones, which can hinder onset detection performance. The use of the logarithmic magnitude has a noticeable effect on the results. Despite the fact that the best performance attained when adding the Mel-scale mapping is almost identical to the one obtained by using only the logarithmic whitening, the ROC curves indicate that the frequency scaling is slightly better when considering a wide range of fixed threshold values. The inspection of the type of errors reveals that most false positives correspond to soft *madera* sounds that are played in between the actual strokes of the pattern as a subdivision, as well as some strokes of the finger tips on the drumhead also played as a subdivision, whereas most false negatives are deaccented stick events.

However, it must be noted that the results of this evaluation can not be generalized straightforwardly to other situations, such as ensemble recordings or other drum types. In order to illustrate this, a separate track of the *piano* drum from one of the ensemble recordings of the dataset introduced in Section 3.4 is considered. The onset detection method previously described was applied, using the same parameter configuration and an arbitrary value for the fixed threshold. The only difference is that, since the *piano* drum is the lowest one, only frequencies up to 500 Hz were summed up. The resulting onsets were manually checked and corrected when necessary, for a total of 612 onsets. Finally, the onset detection was repeated varying the fixed threshold value and changing the different preprocessing steps, namely the logarithmic whitening and the Mel-scale mapping.

The results obtained, which are presented as ROC curves in Fig. 4.8, clearly indicate that the logarithmic magnitude whitening is not appropriate in this case, neither alone nor combined with the Mel-scale mapping. This makes sense, because emphasising high-frequency content when low-frequency onsets are pursued, can only benefit the detection of spurious events from the other interfering drums of the ensemble. Notably, the Mel-scale mapping alone SF^{filt} considerably improves the performance results yielding an F-measure of 0.992.

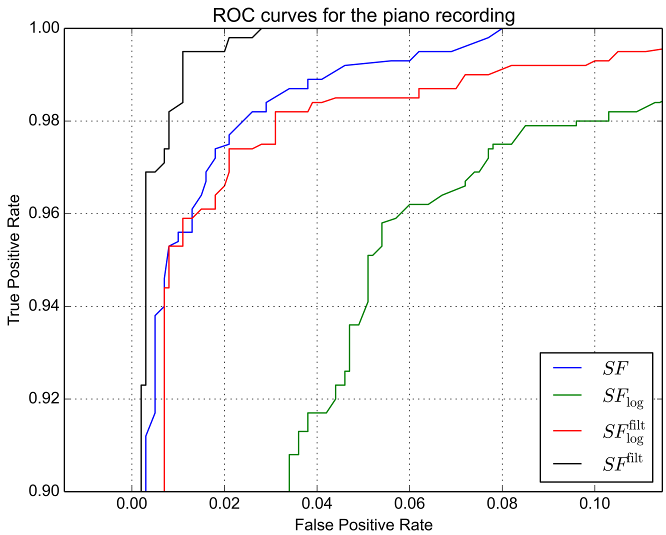


Figure 4.8: ROC curves for the *piano* recording comparing the different detection functions.

4.4 Audio feature analysis and classification

This section reports two types of experiments which involve the classification of detected events, one aiming to recognize the predominant *candombe* drum in an audio file, and the other attempting to identify those temporal segments of a *repique* performance when the *clave* pattern is played.

The classification is addressed by modelling the same audio features used for onset detection, namely the spectral flux. To do that, the STFT of the audio signal is computed in sequential 80-ms duration windows in hops of 20 ms, and then mapped to the Mel-scale using 160 bands. The resulting sequences are time-differentiated (via first-order difference) and half-wave rectified. To produce the onset detection function the obtained feature values are summed along all Mel sub-bands. For drum sound classification, the vector containing the first 40 Mel bands, corresponding to frequencies up to 1000 Hz, is employed. This frequency value was chosen based on some feature selection experiments.

4.4.1 Datasets

A training dataset containing isolated sounds of *candombe* drums was compiled and annotated for this work. To this end, recordings from the preliminary studio session were considered (see Section 3.4.2), in which four percussionists played in turns one among a set of three drums (one of each type) called `drums-1` hereafter. Automatic onset detection was performed over each audio track, and the resulting events were manually checked and labelled as of a certain sound type. A different class was attributed to each drum type (i.e. *chico*, *repique*, *piano*) besides an additional one to *madera* strokes (which sound very similar for all drums). Recording each type of drum separately greatly simplified the manual labelling process, since once *madera* sounds had been identified and labelled in a given track, all remaining events could be assigned to its (known) drum type. Finally, a training dataset of 2000 patterns was built through a stratified random sampling (500 of each class).

Figure 4.9: Testing dataset recording session. Drums on the left are also used for training (`drums-1`), while drums on the right belong to the set used only for testing (`drums-2`).

Another dataset of real performances of drum ensembles was used for testing. This data was collected in the recording session introduced in Section 3.4, in which five renowned *candombe* drummers were recorded in a multi-track audio system, playing in groups of three to five. Two of these group configurations are depicted in Fig. 4.9. Audio recordings were done using spot microphones close to each drum.[4] This provides synchronized audio tracks in which a certain drum is predominant, whilst there is interference from the other drums. Complete performances of variable lengths were recorded, approximately from two to four minutes each. The same set of drums, `drums-1`, used for recording the training samples was used in all three-player performances. Another set of drums, called `drums-2`, was involved in the four- and five-player recordings. This set-up allows for two different types of experiment regarding the generalization ability of the classification system: one in which training and testing drums are the same, but recording conditions (e.g. room acoustics, microphones) and performance configuration (e.g. drum tuning, percussionist) change; and another in which the instruments are also changed.

4.4.2 Clustering and classification methods

In order to explore the training data, a clustering analysis using the K-means algorithm [150] was carried out. The distance measure for the analysis should reflect the similarity in shape between two spectral feature profiles, and turned out to be a key issue since several measures considered were not appropriate.

[4]Except for the *chico* drum in ensembles of five players due to equipment constraints.

	chico	repique	piano	madera
chico	474	23	0	3
repique	78	403	3	16
piano	1	79	420	0
madera	2	2	0	496
%	94.8	80.6	84.0	99.2

Table 4.2: Confusion matrix of a cluster-to-class evaluation for the training data.

The Pearson correlation computed as

$$C(x,y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2}\sqrt{\sum (y_i - \bar{y})^2}} \tag{4.11}$$

$$= \frac{\langle x - \bar{x},\ y - \bar{y} \rangle}{||x - \bar{x}||\ ||y - \bar{y}||} \tag{4.12}$$

corresponds to the inner product of two sequences x and y normalized to zero mean and unit standard deviation, and can be seen as a shift-invariant cosine similarity. By treating the data points as the correlated sequences, their distances can be measured as

$$D(x,y) = \frac{1 - C(x,y)}{2} \ \in [0,1]. \tag{4.13}$$

The component-wise mean of its points is the centroid of each cluster.

The results of this clustering analysis applied to the training data when setting the number of clusters K=4 is presented in Fig. 4.10 and Table 4.2. The confusion matrix in Table 4.2 of a cluster-to-class evaluation, shows that *madera* and *chico* classes are correctly grouped, while *piano* and *repique* exhibit a higher rate of misclassification. A three-dimensional representation computed with multidimensional scaling (MDS) using the same distance measure is included in Fig. 4.10 for data visualization, and highlights the overlapping of classes. In particular, *repique* is the most troublesome class, which is not surprising since this is the drum of medium size and register, and thus expected to overlap the other drums' spectra. This issue is confirmed by the cluster centroids of Fig. 4.10-bottom, whose shape is consistent with the spectral content of each sound class. The centroid of the *piano* drum class has a clear predominance at low frequencies, whereas the centroid of the *madera* class is dominant at high frequencies. At medium frequencies, the centroid of the *repique* class exhibits a maximum towards the lower range, while the centroid of the *chico* class has higher frequency content.

Results of the clustering analysis motivated the idea of testing a very simple classifier based on the obtained centroids: each centroid was considered as a single class prototype in a nearest neighbours classifier (1-NN), using the previously introduced Pearson correlation distance. Such a classification scheme can simplify the process of building the training database, since unsupervised clustering can

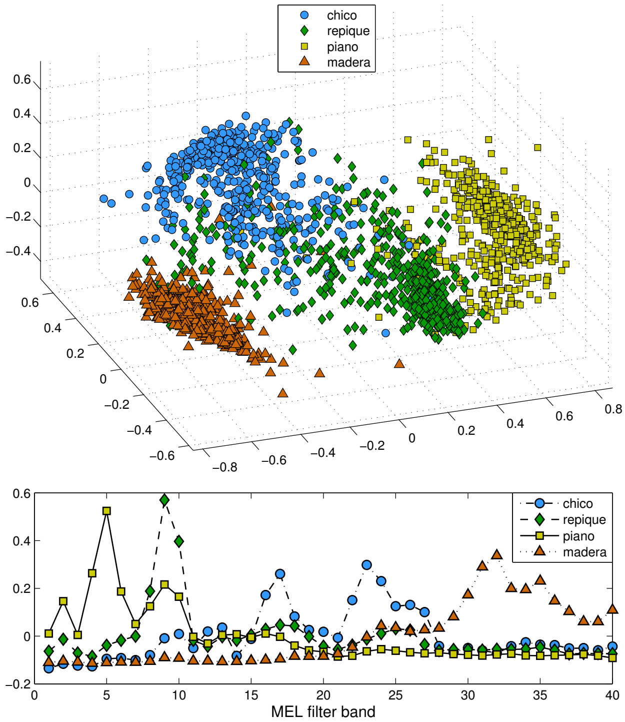


Figure 4.10: Cluster centroids and three-dimensional MDS representation of the training data.

substitute for manual labelling. Furthermore, data coming from different sources, for instance different sets of drums or recording conditions, may be clustered independently so as to better describe classes with more than a single prototype.

Using the same Pearson correlation distance measure, a nearest neighbour classifier (k-NN) and a radial basis function support vector machine (RBF-SVM) were also implemented for comparison. The values for the parameters of the SVM were grid-searched in a cross-validation scheme.

4.4.3 Predominant drum recognition

Recognition of the predominant drum in a given audio track is tackled in a straightforward manner. First, the spectral flux feature is computed, followed by onsets detection, and classification of each detected event into one of the four defined classes. The proportion of onsets in each class gives an indication of the predominant instrument in the audio file.

A simple but effective strategy was adopted to improve the detection of the *repique* drum, already identified in the training phase as the most difficult one. Considering that in a real performance, after the rhythm patterns have been initiated (i.e. after the first few seconds), *madera* sounds are played only by the *repique* drum, the onsets in the *madera* class were included in the *repique* drum class before computing the proportions.

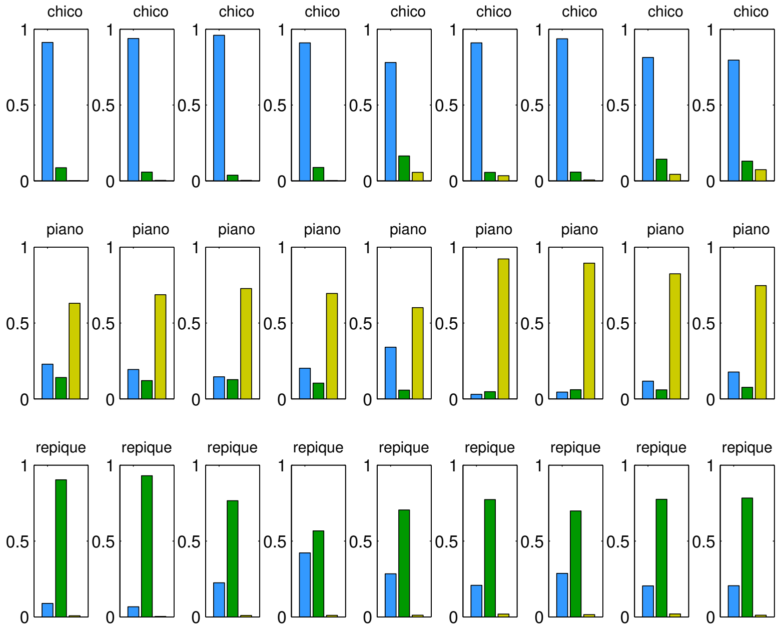
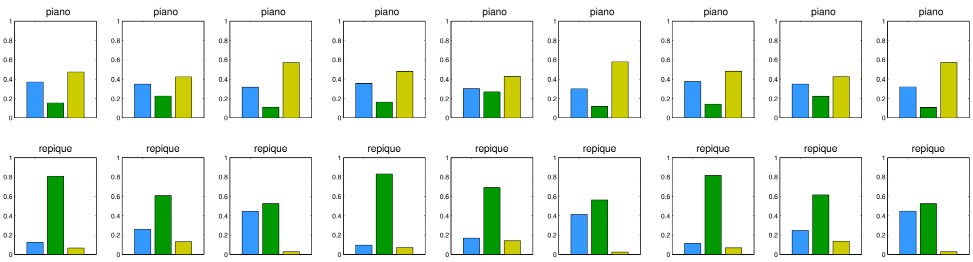


Figure 4.11: Results of predominant drum recognition for the three-drum recordings using a 1-NN classifier of training dataset K-means centroids (■ *chico*, ■ *repique*, ■ *piano*).



(a) 1-NN with centroids (b) 5-NN (c) RBF-SVM

Figure 4.12: Predominant drum recognition for `drums-2` set (■ *chico*, ■ *repique*, ■ *piano*).

In the first experiment set-up all three-drum performances were considered. There are 9 recordings of 3 tracks, totalling 75 minutes and 27 audio files. Note that in this case, the same set of drums of the training samples (`drums-1`) was used. The estimated proportion of onsets for each audio file is shown in Fig. 5.3, for the 1-NN classifier based on the K-means centroid prototypes. It can be seen that the majority class always indicates the predominant drum. Similar results were obtained with k-NN and RBF-SVM, as shown in the next experiments.

The other set of drums (`drums-2`), not used for training, was employed in another experiment. There are 6 different drums, 3 *piano* and 3 *repique* (no *chico*). A track was processed for each drum, totalling 22 minutes of audio. Classification results are presented in Fig. 5.4 for a 1-NN of centroid prototypes, a 5-NN, and an RBF-SVM. Although the majority class always reveals the correct drum type, there is a noticeable difference in the disparity among classes with regards to the previous experiment. This seems to disclose some lack of generalization ability to handle different sets of drums. However, it has to be taken into account that these recordings involve more than three drums, which reduces the distance between performers (as seen in Fig. 4.9) and therefore increases the interference (e.g. *chico* in the *piano* tracks for five-player recordings). Differences among classifiers are marginal, and results are very similar for different choices of k-NN neighbours.

4.4.4 Detection of *clave* pattern sections

A similar approach was followed for detecting those sections when a *repique* drum plays the *clave* pattern. Five performances in which two *repique* drums take part were chosen for this experiment, totalling 10 tracks and 33 minutes of audio.

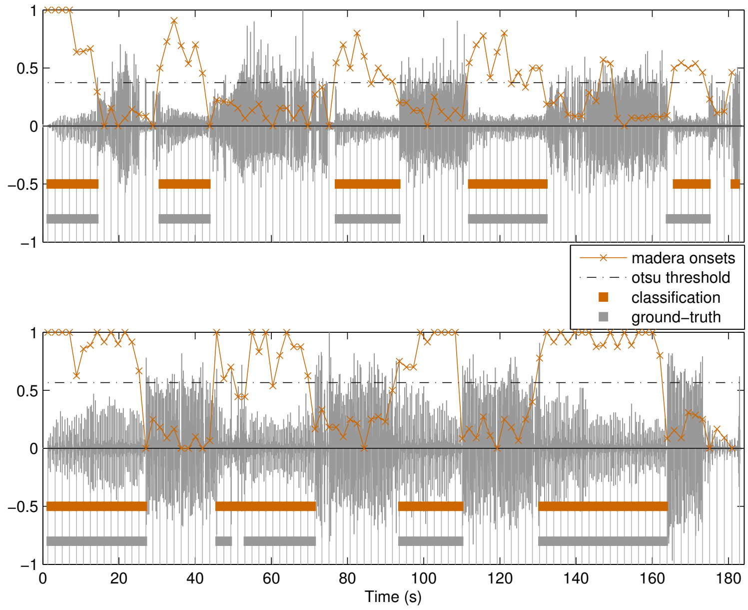
A *clave* pattern lasts for a whole musical bar; therefore, the recordings were manually labelled indicating all bar locations as well as which of them contained the *clave* pattern. The onsets in each track were detected and classified. Then, the proportion of *madera* onsets to the total detected events within each bar was computed as an indication of the presence of the *clave* pattern. A two-state classification was performed according to a threshold computed using Otsu's method [226]. Finally, to avoid spurious transitions, a hysteresis post-processing was implemented in which a change of state is validated only if it is confirmed by the following two points of the sequence. The segmentation process is illustrated in Fig. 4.13-left for two of the audio tracks.

The performance error attained by the three classifier schemes for each audio track, computed as the percentage of bars in which annotation and classification are different, is presented in Fig. 4.13-right.

4.5 Discussion and conclusions

This chapter introduced the audio features used along the thesis, following a typical approach based on the spectral flux, which is well-suited for dealing with the percussive events of *candombe* drumming. Besides, a sub-band analysis which provides a rough separation of the rhythmic patterns by exploiting their different registers was described. Moreover, a representation in the form of a map of features was proposed to study the differences and similarities of the rhythmic patterns, as well as their evolution throughout a performance.

Then, the features were assessed when applied to the onset detection and classification of *candombe* drum sound events. They show to be effective for onset detection, though the optimal configuration may depend on the particular type of drum at hand and of the problem addressed. An approach for predominant drum recognition in audio signals was described, based on modelling the same spectral flux features used for onset detection. The reported experiments yielded promising results, even for the 1-NN classifier of centroid prototypes. To this regard, the Pearson correlation measure—which captures the similarity in shape between two



Classification error (%) for each track		
1-NN	5-NN	SVM
8.4	5.9	8.4
1.7	5.4	4.2
2.0	1.0	3.1
2.0	2.0	2.0
9.3	6.5	6.5
1.0	1.0	1.0
2.7	1.4	2.7
2.7	1.4	2.7
0.8	4.7	0.8
11.6	10.9	10.9
4.2	4.0	4.2

Figure 4.13: Detection of *clave* pattern for two *repique* tracks of the same performance (left) and classification error for each track of the dataset (right). For each waveform plot: in the upper part, the proportion of *madera* onsets detected within each bar is depicted along with the Otsu threshold; in the lower part, vertical lines indicate the labelled bars, while horizontal thick lines show classification and ground-truth labels.

spectral profiles—plays an essential role, which should be further assessed in our future work.

Automatically detecting *clave* patterns from audio recordings was also tackled in this chapter and can be a valuable tool for studying performance in musicological research. For instance, the interaction of two *repique* drums playing together is clearly visible in Fig. 4.13. Sections in which a performer plays the *clave* pattern show an almost perfect anti-symmetry between the two tracks. Besides, there exist several variations of the *clave* pattern that deserve a thorough study, as shown in Section 5.2. To do that, the automatic detection of *clave* sections in a recording allows for dealing with large audio collections. In addition, *clave* pattern serves as a mean of temporal synchronization and could be exploited by automatic rhythm analysis algorithms for beat and downbeat tracking.

Chapter 5

Analysis of rhythmic patterns

This chapter is based on work originally reported in [254], [159], [160] and [251]. The description herein reproduces some passages of the papers and includes modifications and additions in order to put the work in the context of this dissertation.

5.1 Introduction

There is a broad agreement on the importance of rhythmic patterns as structural elements in music [193]. From Western Africa traditions to European folk dances, repetitive rhythmic patterns are at the core of the rhythmic/metrical structure. The study of rhythm has a long tradition in music theory and musicology. Its structure is often regarded as a hierarchy of different levels, which is inferred by the listener through a complex cognitive process [181].

In recent decades, empirical music studies have applied computational approaches to deal with symbolic music, using software tools such as the Humdrum Toolkit [146] and music21 [79]. At the same time, research in music information retrieval (MIR) has undertaken the development of techniques for extracting musically meaningful content information from the automatic analysis of data collections, such as audio recordings or symbolic music. In this context, there is a lot of work on the characterization of repetitive patterns to address topics like music structure and similarity [35]. Part of this research deals specifically with rhythmic patterns. For instance, bar-length drum patterns computed from symbolic music have been used for studying musical rhythm by the application of statistical methods from natural language processing [200]. In some other works, rhythmic patterns are automatically extracted from the audio signal. For example, bar-length rhythmic patterns computed from the energy evolution of the audio signal have been applied to the characterization of music and genre classification [95]. Recently, the explicit modelling of rhythmic patterns has been proposed as a way to improve upon existing beat-tracking algorithms, which typically fail on dealing with syncopated or polyrhythmic music [174]. Those rhythmic patterns, that describe the distribution of note onsets within a predefined time interval, can be learned from audio signals, thus enabling the model to adapt to any kind of music.

The analysis of micro–rhythmic aspects of music has received an increasing amount of attention in recent years, and has developed a more solid theoretical framework [44, 45, 144]. Micro–timing involves small–scale temporal deviations of events in the musical surface with respect to an underlying regular metrical grid. The systematic use of these deviations can be of structural importance in the rhythmic and stylistic configuration of some genres. In some cases, these deviations take the form of tempo variations like *rubato*, *accelerando* or *ritardando*; this is common practice in traditional Western art music from Baroque to Romanticism [144,145]. In other contexts, however, micro–timing is more appropriately represented by the time–shifting of events with respect to the steady beats of a constant tempo, e.g. *notes inégales* in Baroque, or "swing" eighth–notes in Jazz [38, 63, 92, 209]. This practice is an important characteristic of many genres of contemporary popular music and in some traditional musics of the Afro–Atlantic culture [82,112,149,216]. It has also been argued that in some cases micro–timing could be better understood when integrated into the metrical framework by considering non–isochronous beat subdivisions [239].

In this chapter some techniques are proposed for the detailed analysis of the rhythmic patterns of *candombe* drumming from audio recordings. As noted in Section 2.3.2, each type of drum has a distinctive rhythmic pattern and a timeline pattern, called *madera* or *clave*, is shared by all drums. Some of these rhythmic patterns exhibit different variations and possible ornamentations. This is the case of the *piano* drum, whose rhythmic patterns are considered to be stylistic markers. Similarly, the *clave* pattern in *candombe* allows for several different types and variants. On the other hand, some of the patterns of the rhythm show virtually no change. This is the case of the *chico* drum, which plays an ostinato of sixteenth notes throughout the whole performance. However, it has been suggested recently that this pattern exhibits characteristic micro-temporal deviations [112], which deserve a thorough study. Likewise, it has been noted that the *repique's* primary pattern shows a deviation with respect to the four pulses of the beat [157].

Therefore, a set of tools is proposed for the study of rhythmic patterns that span over the four-beat cycle, with the aim of investigating its different types and forms. An interactive software is implemented, to allow for the analysis of the rhythmic patterns in a given recording, and reveals different aspects of the performance. First, a set of experiments is presented which focus on the analysis of the *piano* drum from recordings of ensemble performances. A data-driven approach, applied to annotated audio signals, yields characteristic patterns of the instrument and allows the study of differences and similarities among performance styles. In turn, the results of the *piano* pattern analysis presented herein are useful to inform the proposed scheme for supervised rhythmic/metrical tracking described in Chapter 6. Then, the same techniques are applied to the rhythm cycles where the *clave* pattern is played in a given recording. The analysis shows the different *clave* pattern variants used by a single player throughout a performance and how they are alternated. After that, all the rhythm cycles collected from a dataset of recordings are aggregated, and the different ways in which the *clave* pattern is played are studied. Additionally, other types of experiments are proposed in order to assess the exact nature of the micro-temporal deviations found in the rhythmic cells of the *chico* and *repique* patterns. The analysis of several recordings by renowned players reveals the systematic and consistent use of micro–temporal deviations, suggesting that micro–timing is a structural component of *candombe* rhythm.

5.2 Rhythmic pattern analysis

This section focuses on the analysis of rhythmic patterns that span over the whole rhythm cycle. Firstly, the analysis methods applied and the software tools implemented are described. Then, some experiments that deal with the *piano* drum and *clave* patterns are presented.

5.2.1 Analysis methods

Feature extraction

The audio recording is processed to extract the spectral features described in Section 4.2, as summarized in the following. The STFT of the audio signal is computed in sequential 40-ms duration frames in hops of 10 ms, weighted by a Hann window, and then mapped to the Mel-scale using 160 bands. No logarithmic magnitude compression is applied. The resulting sequences are time-differentiated (via first-order difference) and half-wave rectified. A sub-band analysis is performed to focus on the corresponding frequency bands of the pattern of interest. In particular, the low-frequency bands up to approximately 200 Hz are summed for the analysis of the *piano* drum, while the high-frequency bands from approximately 1000 to 1800 Hz are summed for the study of the *clave* pattern. Then, by using the beat/downbeat annotations of the recording, the features are normalized by the 8-norm within a local window of half a rhythm cycle, time-quantized to a grid of *tatum* pulses, and grouped into cycle-length vectors. Hence, the performance can be represented as a map of cycle-length rhythmic patterns, as shown in Figure 5.1.

Unsupervised clustering

Cycle-length rhythmic patterns are unsupervisedly clustered to aid the analysis of their differences and similarities. Different techniques can be applied for this task [150]. Among them, the classical K-means method and spectral clustering were selected. The K-means method is an iterative algorithm which aims to partition the data patterns so as to minimize the within-cluster sum of the distances between patterns and their mean. It starts with a randomly selected set of cluster centroids and partitions the data patterns by assigning each one to its closest centroid. Then, new centroids are computed as the mean of the patterns within each cluster and the data is partitioned again. The process is repeated until the pattern-to-cluster assignments no longer change. The obtained centroids serve as prototypes of the clusters. The main drawback of the method is that it can not handle non-convex clusters properly.

Conversely, spectral clustering is useful when the clusters are non-convex or are not suitably described by a measure of centre and spread. The method does a low-dimension embedding of the patterns, given by the Laplacian of their similarity graph, followed by a K-means clustering in the low dimensional space [292]. Building the similarity graph is not a trivial task and is a key factor in spectral clustering performance. Different alternatives exist for the type of graph, such as k-nearest neighbour, ϵ-neighbourhood or fully connected graphs, which behave rather differently. Unfortunately, barely any theoretical results are known to guide this choice and to select graph parameters [292]. A general criteria is that the resulting graph should be fully connected or at least should contain significantly fewer connected components than the clusters we want to detect. Otherwise, the algorithm will trivially return connected components as clusters. In the experiments conducted a k-nearest neighbour graph was adopted.

The number of clusters K has to be specified as an input parameter for both algorithms. Even if several automatic strategies do exist to address the estimation of the number of clusters, none of them is without fail, so a manual selection relying on the visual and aural inspection of the resulting clusters was favoured when carrying out the experiments. The automatic selection of the number of clusters (i.e. different rhythmic patterns) in a recording is addressed in Chapter 7 using information theory concepts. With regard to the definition of similarity between patterns, Euclidean distance and cosine similarity were considered, both yielding very similar results, probably due to the normalization of feature values. Therefore, the former is used in the reported experiments.

Both clustering algorithms proved to be effective for the problem at hand, usually providing equivalent solutions. Thus, considering that, apart from specifying the number of clusters K, the spectral clustering method requires the selection of the graph parameters, the K-means algorithm was applied as the default method in the reported experiments.

Dimensionality reduction and manifold learning

For visualization purposes, the patterns are mapped to a low dimensional space. There are several approaches for dimensionality reduction [261], among which multi-dimensional scaling (MDS), isometric mapping (Isomap), locally linear embedding (LLE), and spectral embedding were applied. Given that the main aim of this processing is visualizing differences and similarities among rhythmic patterns, MDS and Isomap were preferred, since they are capable of keeping the levels of similarity among the original patterns after being mapped to the lower dimensional space. In contrast, the two remaining techniques produced less meaningful representations in the conducted experiments.

Metric MDS is based on computing the low dimensional representation that most faithfully preserves the pairwise distance between input patterns. The input to MDS is specified as a matrix of pairwise distance between patterns, from which a Gram matrix is derived [261]. Then, the solution is obtained from the spectral decomposition of the Gram matrix. The Isomap method extends MDS by taking into account the intrinsic geometry of the data manifold through an estimate of the

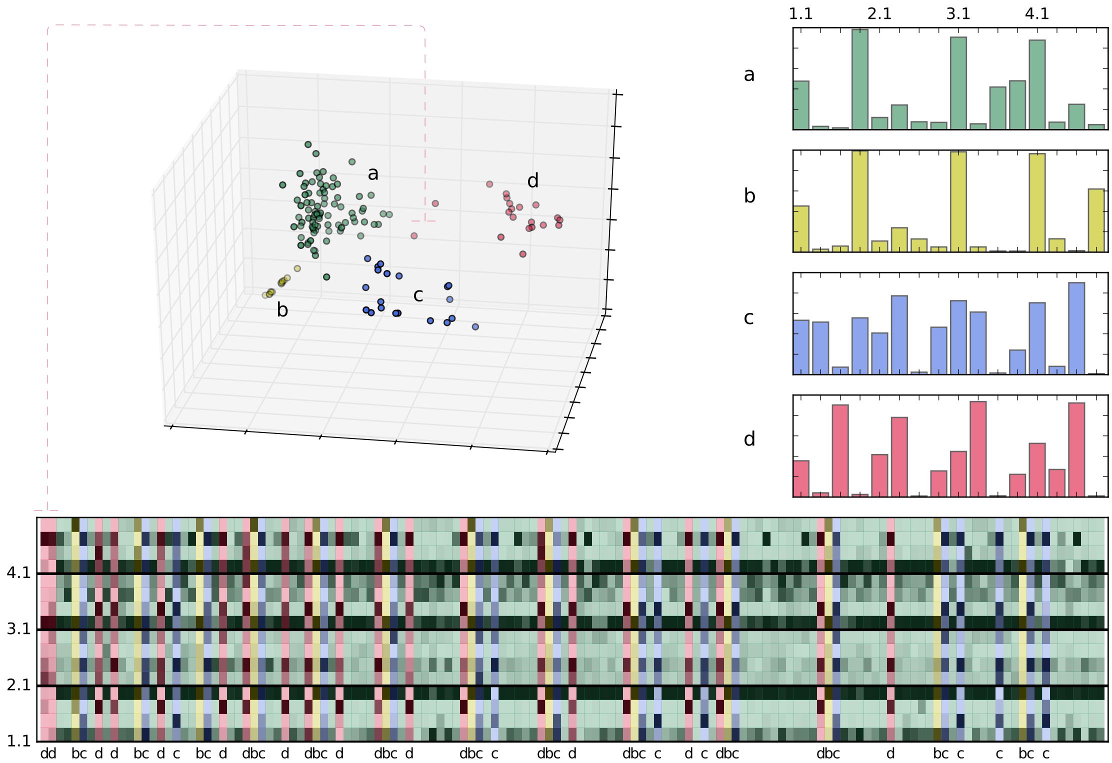


Figure 5.1: Clustering analysis of *piano* patterns in recording of Exp. 4.1. The three-dimension representation is computed using Isomap. Centroids of clusters are shown at the right. Letters at the bottom of the map indicate the assigned cluster (cluster a omitted for clarity).

geodesic distance between patterns, which is obtained from a neighbourhood graph. The estimated geodesic distances are then used as input to the classical MDS algorithm. For this reason, Isomap is more effective for analysing data structures that lie in a manifold of the original space, and was the method applied in the reported experiments. Besides, it allows the projection of new patterns onto the low-dimensional space derived from the training patterns. The nearest neighbours of the new patterns in the training data are found, and from them the shortest geodesic distances to each point in the training data are computed. This was exploited in the third experiment of Sec. 5.2.2 in order to map new patterns onto a low dimensional space obtained from training data patterns.

5.2.2 Experiments and results

The above described tools were implemented in an interactive software that allows the analysis of the existing patterns in a *candombe* recording. The system loads an audio file together with its corresponding beat/downbeat labels, and produces a cycle-length map of rhythmic patterns and a clustering analysis. The dataset of recordings described in Section 3.3 was used in a set of experiments intended for the study of *piano* drum patterns, part of which are reported in the following.

Analysis of piano drum patterns in a recording

Experiment 5.1 An example of the analysis of the *piano* drum patterns in a recording is shown in Fig. 5.1. The recording, a little over 4 minutes long, features noted and influential *piano* player Gustavo Oviedo, in a group comprising also a *chico* and two *repique* drums, all of them belonging to the *Ansina* style. The first remark about the results is that the obtained clusters match characteristic patterns of the instrument. In this case, the most frequent pattern is a typical *base* pattern (cluster *a*, green), while the other clusters correspond to an alternate *base* pattern (cluster *b*, yellow), and to two different types of *repicado* patterns (clusters *c* and *d*, blue and red respectively). The first two patterns of the recording are readily seen as outliers of cluster *d* (red) in the low-dimensional representation, marked with a dashed red line. In these first rhythm cycles the *piano* drum plays the *clave*, so the patterns are meaningless in terms of *piano* patterns. Fortunately, this sort of situation can be easily spotted by means of the implemented software, which allows one to listen to each individual pattern. The map of bar-length patterns also permits an easy visualisation of important structural aspects of the recording, like the irregular but well-balanced distribution of *base* and *repicado* patterns along its duration. It also reveals that the alternate *base* pattern, i.e. cluster *b* (yellow), is always followed by a *repicado* pattern belonging to cluster *c* (blue).

Fig. 5.2 shows the patterns corresponding to each cluster in symbolic music notation.[1] For the sake of clarity of notation, all the variety of different strokes and articulations that can appear in the performance of a skilled player has been reduced to six basic typologies: three with the hand (open, muffled, and finger tips, indicated with a triangular note head), and three with the stick (open, muffled and a press roll or buzz, indicated with a trill symbol). The muffled tones are indicated with a cross. Comparing Fig. 5.2*a* and *b*, it can be seen that they begin in a similar fashion, but in the alternate *base* pattern the last beat introduces an anacrusis of the *repicado* in Fig. 5.2*c*. Fig. 5.2 also shows variants of the primary *base* and *repicado* patterns, which exhibit differences only in the last beat.

Comparison of piano drum performance styles

In order to compare *piano* drum performance styles, only the *base* patterns were considered. Notwithstanding, it is worth noting that other types of comparison can be carried out, for instance, by also including *repicado* patterns and by considering their temporal evolution. First of all, the clustering analysis of a recording was conducted as described above and only the patterns within the largest cluster were selected for further processing. This is based on the hypothesis that the *base* pattern is performed most of the time. Anyway, the selected cluster was aurally inspected using the implemented software to assess whether it was representative enough. This is an interactive process that may involve choosing different values for the number of clusters (parameter K), until an appropriate configuration is selected. Then, *base* patterns are grouped and classified into separate classes that tend to match different performance styles, as shown in the following experiments.

[1]Music notation of these patterns is based on the one provided by Luis Jure in [254].

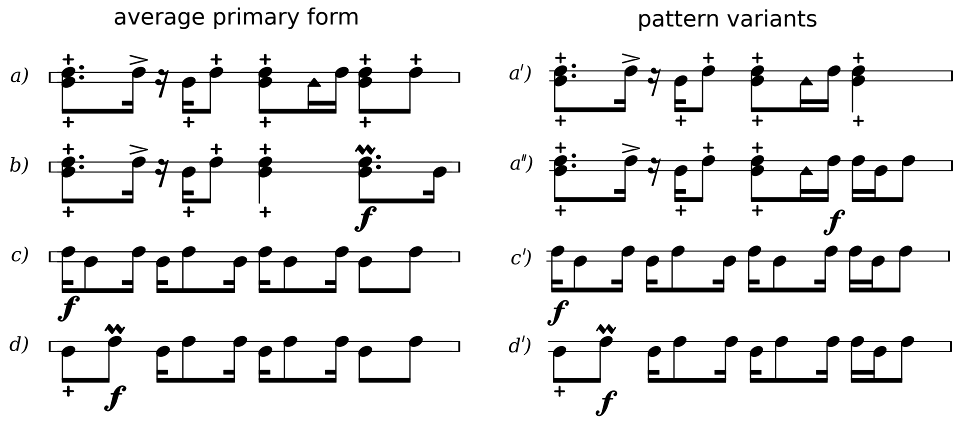
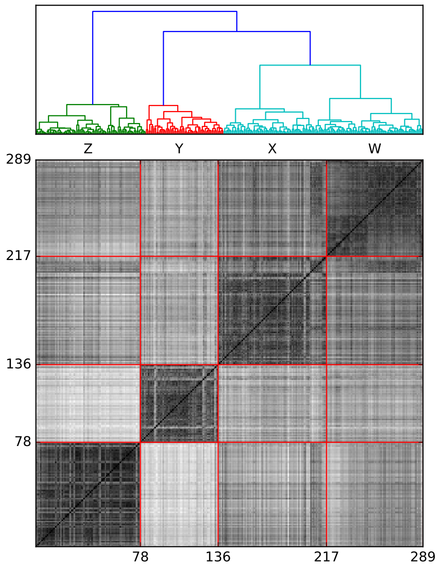


Figure 5.2: Left column shows the patterns in their average primary form: *a)* primary *base* pattern (green); *b)* alternate *base* pattern, with the *repicado* beginning on the fourth beat (yellow); *c)* main *repicado* pattern (blue); *d)* alternate *repicado* pattern (red). Right column shows variants of the primary *base* and *repicado* patterns, which differ only in the last beat.

Experiment 5.2 In the first experiment a set of four recordings was considered, containing performances by different groups of players of diverse styles. The first recording is from the style of the Cordón neighbourhood and is played by Rodolfo "Pelado" Rodríguez, while the second one is from the virtuoso *piano* player Eduardo "Malumba" Giménez, who belongs to the Ansina tradition. The remainder two recordings are both from a third style, namely Cuareim, and performers are Fernando "Lobo" Nuñez and Luis "Zorro" Pereira.

The similarity matrix of patterns sorted by performer is depicted in Fig. 5.3-left. It has a block-diagonal shape which reveals the similarity between patterns of the same performer. Following the order in which the recordings were introduced, the number of *base* patterns in each of them is 78, 58, 81 and 72 respectively, and players are labelled as Z, Y, X and W. It can also be seen that the patterns of the last two players tend to be more similar, which is probably related to their common traditional style. These remarks are also consistent with a hierarchical cluster analysis using Ward's linkage method [162], which is shown in the top of the same figure. Despite their similarities, the patterns of each performer are clearly separable in a two-dimensional representation computed using the Isomap method, as shown in Figure 5.3-right, where decision boundaries of a k-nearest-neighbour classifier are represented. It is important to notice that the decision boundaries are not much sensitive to the number of neighbours k chosen for classification.

Experiment 5.3 The results of the previous experiment suggest that the *base piano* patterns not only reveal performance styles in a broad sense but are somehow distinct for each different performer. Therefore, in the second experiment this was further explored by considering two different recordings of each performer, in order to assess to what extent the patterns in one recording resembled the patterns in the other. The recordings involve four different performers, namely Eduardo "Cacho"



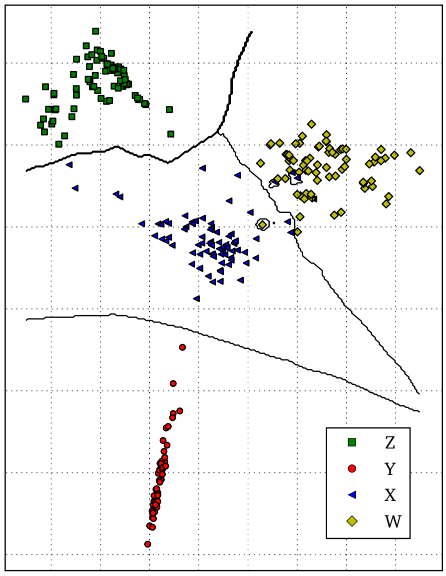


Figure 5.3: Results for Exp. 4.2. Left: Similarity matrix of the *piano* base patterns sorted by performer (bottom) and dendrogram of the hierarchical clustering analysis (top). Right: Comparison of *piano* drum patterns for different players in a two-dimensional space computed using Isomap. Decision boundaries for a k-nearest-neighbour classifier are depicted (k=7).

Giménez and Gustavo Oviedo from Ansina, and Waldemar "Cachila" Silva and Juan Silva from Cuareim.

Fig. 5.4 shows the similarity matrix for the patterns in this dataset. The performers are labelled in the above order from A to D, and the number of patterns in the corresponding pairs of recordings are 150/133, 85/93, 164/140 and 185/188 respectively. As in the previous experiment, the block-diagonal shape of the similarity matrix is also visible, but in addition a secondary diagonal can be discerned, which discloses the similarity of patterns of the same performer in a different recording. This was further evaluated by building a k-nearest-neighbour scheme with the patterns in one recording of each performer as the train set, and then classifying the patterns in the remaining audio files as the test set. Additionally, the same procedure was applied to the patterns mapped to a three-dimensional space using the Isomap method. The obtained results, which are presented in Table 5.1, show that classification accuracy is far beyond random choice rate (25%), even in the three-dimensional space. This seems to emphasize the ability of *base* patterns of the *piano* drum to describe personal styles. It is interesting to note that the highest confusion rate takes place between the last two players, which not only share the same traditional style but are also brothers.

Analysis of clave patterns

The techniques were also applied to all the ensemble performances from the dataset described in Section 3.4 so as to analyse the characteristics of the *clave* pattern, and

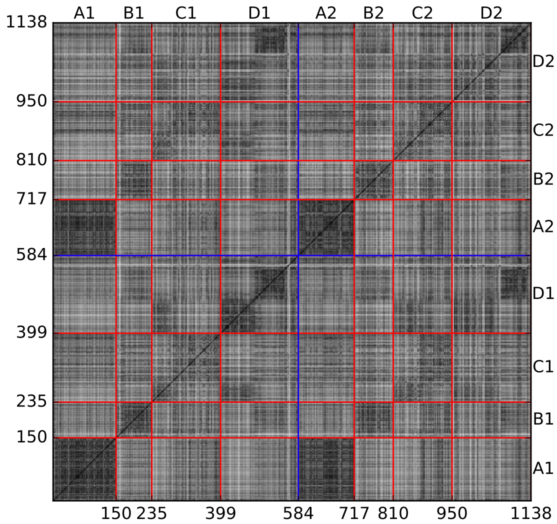


Figure 5.4: Similarity matrix for the two recording sets of Exp. 4.3. A secondary diagonal can be discerned, which discloses the similarity of patterns of the same performer in different files.

	original dimension, 16				**lower dimension, 3**			
	A	B	C	D	A	B	C	D
A	126	1	3	3	116	2	9	6
B	2	71	12	8	2	69	13	9
C	7	6	106	21	8	11	87	34
D	7	3	35	143	8	13	55	112
%	94.7	76.3	75.7	76.1	87.2	74.2	62.1	59.6

Table 5.1: Confusion matrix and classification performance for a k-nearest-neighbour classifier (k=7), in the original space and in a three-dimension mapping computed using Isomap.

ascertain how it is integrated in the rhythm. To do that, the separate audio tracks of the *repique* drums were considered. This is because, after the introduction to the rhythm—in which all drums play the *clave* pattern—, it is only the *repique* drum that plays *madera* in between phrases. The rhythm cycles in a given recording where the *clave* pattern is played can be faithfully identified using automatic onset detection and sound classification based on spectral timbre features, as described in Section 4.4.4. Yet, the experiments reported herein relied on manual annotation, to avoid any spurious result due to automatic classification errors. The analysis of an audio track provides information on the different patterns used by a single player and how they are alternated throughout the performance. The first experiment is an example of this type of analysis. In the second experiment, all *clave* rhythm cycles collected from the dataset were clustered according to their similarity. Each group represents a different way of playing the *clave*, which were analysed and compared to the ones reported in the musicological literature.

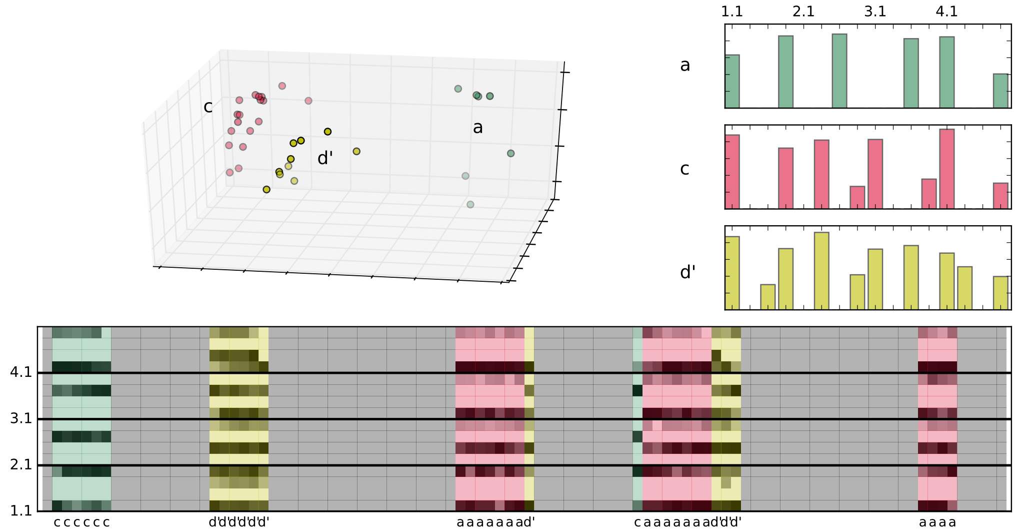


Figure 5.5: Clustering analysis of *clave* patterns in recording of Exp. 4.4. The representation in three dimensions is computed using Isomap. Centroids of clusters are shown at the right.

Analysis of clave patterns in a recording

Experiment 5.4 An example of the analysis of the *clave* patterns in a recording of the dataset is shown in Fig. 5.5. The recording is about 3 minutes long and features an ensemble of four drums. The audio track analysed corresponds to a *repique* drum played by Sergio Ortuño. The ensemble also includes a *chico*, a *piano* and another *repique* drum. Following the introduction to the rhythm, in which all the performers play the *clave* pattern, the two *repique* drums take turns to improvise while the other plays *madera*. This is actually the same recording of Fig. 4.13, where the interaction between the *repique* drums is clearly visible.

With regard to the results, the first remark is that the player made use of three different kinds of *clave* patterns during the performance—referred to as *a*, *c* and *d'*, to be consistent with the notation to be used in Exp. 5.5—whose temporal location tends to be organized into sections of the same type. The distinction between the pattern types is consistently disclosed within the second and third beats, in particular through the differences at subdivisions 2.2–2.3 and 3.1–3.3. The performance begins with pattern *a*, which is essentially the five–note syncopated pattern already introduced in Fig. 2.12 as a prototype of the *clave*, and also corresponds to the Cuban 3:2 son *clave*. At present, this is arguably the most common and widespread *clave* pattern of *candombe* rhythm [88, 106, 124, 158]. It was the pattern played most frequently at the beginning of the performance in the recordings analysed (11 out of 14 started with this pattern, the others beginning with pattern d). Still, it only represents about one fifth of the total number of the *clave* rhythm cycles in the dataset, as shown in the following experiment.

Analysis of clave patterns in the dataset

Experiment 5.5 A total of 423 rhythm cycles identified as *clave* patterns were gathered from the different recordings of the dataset. A cluster analysis was carried out with the whole set of *clave* patterns, whose results are shown in Fig. 5.6. In the three-dimensional representation it can be readily seen that there are essentially four main groups, two of which can be subdivided into two variations—denoted as *a*, *b*, *b'*, *c*, *d* and *d'*. All the collected patterns, sorted by cluster, are depicted in the map at the bottom of Fig. 5.6, and the percentage of each type is also indicated. The centroids of the clusters are illustrated as well. The patterns in their average primary form are represented in music notation in Fig. 5.7.[2]

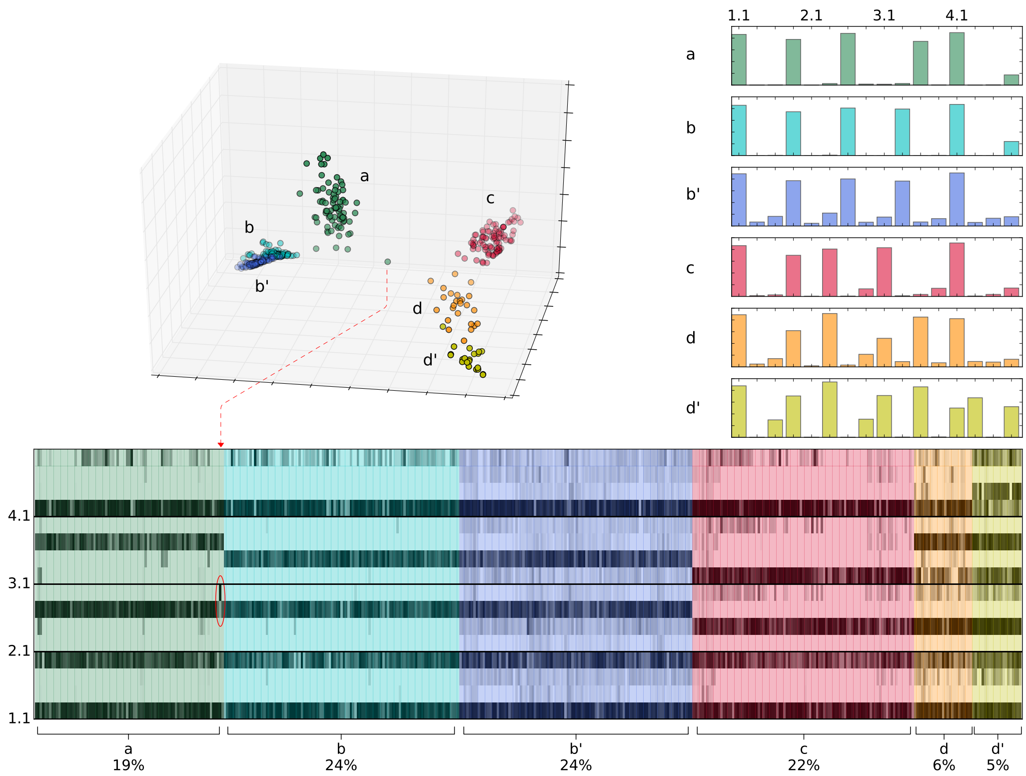


Figure 5.6: Clustering analysis of the *clave* patterns in the dataset. The three-dimension representation is computed using Isomap. Centroids of clusters are shown at the right.

The inspection of the different representations of the patterns allows for an analysis of their similarities and differences. Pattern *a*, already introduced, is described in the musicological literature as the most usual or common pattern, and the one always used to clap along to *candombe* music [106,124,158]. It divides the rhythm cycle irregularly with only two strokes out of five coinciding with a beat. It is worth noting that the articulation of the last subdivision, i.e. 4.4, is an ornamental device that can actually appear or not in any *clave* pattern,

[2]Music notation of these patterns is based on the one provided by Luis Jure in [158].

functioning as an anacrusis of the downbeat, rather than a fundamental stroke. If the ornamentation at 4.4 is not taken into account, nor the weak embellishments such as the ones in patterns *c* and *d*, a representation of the skeleton of each *clave* pattern can be built. This is shown in Fig. 5.8 using a circular rendering of the rhythm cycle, in which consecutive note locations are connected to form a convex polygon. Such a representation enhances the visualization of certain traits and has been used to analyse rhythmic structures [284, 285].

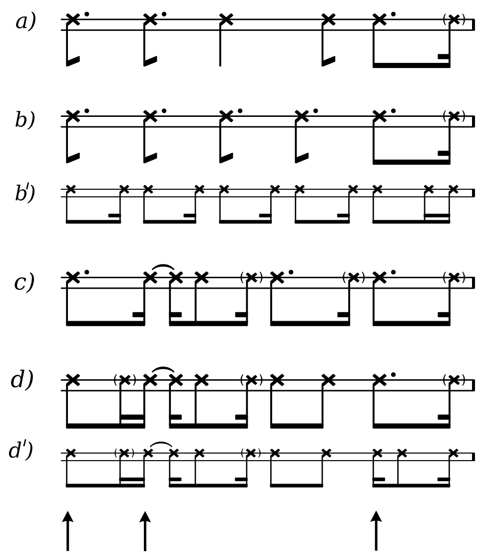


Figure 5.7: The *clave* patterns derived from the dataset in their average primary form represented in music notation. Arrows indicate the three points in which all *clave* patterns coincide.

Turning now to pattern *b*, it is a *clave* pattern that has also been reported in the musicological literature [106, 158], and is sometimes linked to the Ansina style. Compared to pattern *a*, it corresponds to displacing the stroke at subdivision 3.3 to 3.2, as indicated with an arrow in Fig. 5.8-a. The swap of this stroke contributes to the evenness of the pattern, because it results into four out of five inter-onset intervals of the same length. Nevertheless, the pattern still articulates only two beats. Actually, among the five stroke patterns it is a maximally-even set, i.e. it has its elements as evenly spaced as possible [285]. And if the ornamentation at 4.4 is added, it is also a maximally-even set among the six stroke patterns. Note that *b'* is a saturated version of pattern *b*, that softly articulates every subdivision before a stroke of the latter. As a whole, pattern *b* accounts for almost half of the *clave* rhythm cycles in the dataset. The fact that the group of performers featuring in the recordings belong to the Ansina tradition reinforces the idea that the pattern is somehow typical of the style.

Then, pattern *c* can be regarded as the kernel of a different type of *clave* pattern. Compared to pattern *b*, it involves two stroke swaps, i.e. displacing 2.3 to 2.2 and 3.2 to 3.1, as indicated with arrows in Fig. 5.8-b. Hence, the resulting pattern is less syncopated, articulating three consecutive beats. Note that pattern *c* can be ornamented with a soft articulation before each stroke coinciding with a beat, a device that is also reported in the literature [106]. Finally, pattern *d* makes use of the same strokes as the previous pattern but reintroduces an articulation in subdivision 3.3, as in *clave* pattern *a*, which is depicted with an arrow in Fig. 5.8-c. Consequently, the resulting pattern *d* has six definite strokes, and also allows for some embellishments through soft articulations at 1.3 and 2.4. The variant of this pattern, denoted as *d'*, differs in the insertion of a distinctive stroke at 4.2, and, albeit not so relevant, a tendency to articulate subdivision 4.4. It is noteworthy the resemblance of this saturated 10-stroke pattern *d'* to the Afro-Cuban *cáscara* pattern [39]. The patterns *c*, *d*, and its variants are sometimes referred to as a 'traditional' or 'old' *clave* patterns [88], based on the idea that they were more common in the past. Both patterns taken together account for about one third of the *clave* rhythm cycles in the dataset. For completeness, one may note that two simple operations allow to transform back from pattern *d* to pattern *a*, namely the swap of subdivision 2.2 to 2.3 and the deletion of the stroke at 3.1. This could be formalized by defining a measure of the minimum number of basic operations needed to transform one pattern into the other, as in [285].

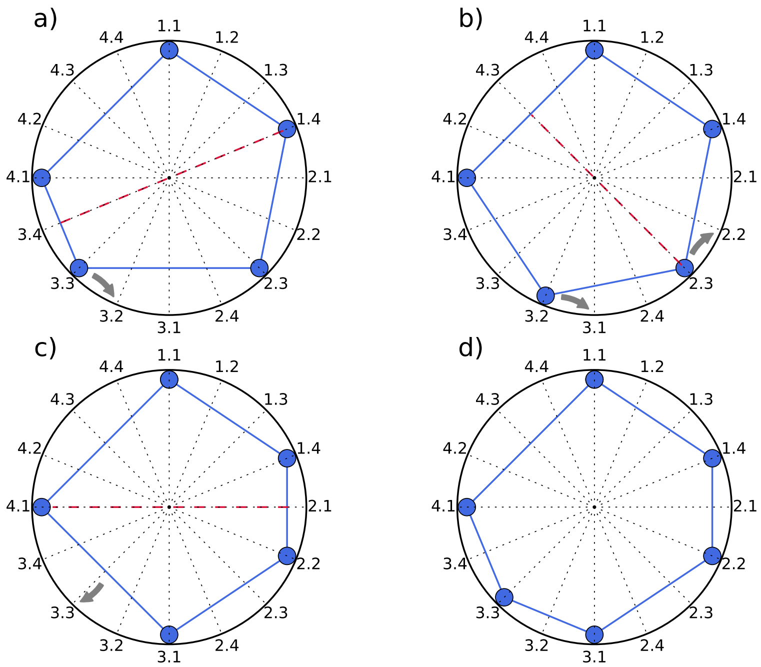


Figure 5.8: Rhythmic patterns in polar notation.

It is worth noting that the cluster analysis shows an outlier, which can be clearly identified in the centre of the three-dimensional representation of Fig. 5.6—marked with a dashed red line to ease the visualization—and is assigned to pattern *a* cluster. The inspection of this pattern reveals that it has a fundamental difference compared to pattern *a*, given that it displaces the articulation at subdivision 2.3 to 2.4, resulting in a rumba—instead of a son—*clave* pattern. Actually, this is also a typical *candombe clave* pattern reported in the literature [106, 158], which has already appeared in the analysis of a field recording shown in Fig.4.4. If a larger set of recordings had been analysed, this pattern would probably have emerged as a well-defined cluster by itself. Including this last case, the analysis yielded virtually all the *candombe clave* patterns reported in the literature [106, 158].[3]

The representation of the patterns as convex polygons illustrated in Fig. 5.8 uncovers certain properties useful for geometric analysis and comparison [284]. For instance, an isoceles triangle indicates two equal consecutive time intervals between strokes. From this perspective, it is immediately obvious that patterns *a* and *b* share the same evenness within the first two beats, given by the articulations at 1.1, 1.4 and 2.3. Likewise, the three isoceles triangles in pattern *b* give account of its maximally evenness. On the other hand, an axis of mirror symmetry indicates that there exists a position from which the pattern sounds the same when played forward or backwards [284]. In this case, patterns *a*, *b* and *c* exhibit such an axis of mirror symmetry—depicted with a red dashed line in Fig. 5.8—that reveals they are *weak palindromes*, since the symmetry is not about the downbeat.

More importantly, the representation also discloses that the triangle defined by subdivisions 1.1, 1.4 and 4.1 is a trait in common among all the *clave* patterns. Indeed, the several possible variants of the *clave* pattern all coincide in the these three points, marked with arrows in Fig. 5.7 [158]. Moreover, the different *base* patterns of the *piano* drum consistently articulate them; see for instance Fig. 5.2 or the survey of patterns reported in [106]. For this reason, it has been suggested that these three points constitute the pillars of the *candombe* rhythmic structure [158]. The first one coincides with the first beat of the bar and is the primary metric accent. As such, it is the point of resolution and conclusion of the rhythm. The other two points are the structural accents of the musical phrase. The one that precedes the second beat of the bar, is a stress accent and represents a "bottom up" rhythmic energy load, while the second one, coinciding with the fourth beat of the measure, is an unload accent. The location of these two accents, and the fact that none of them are coincident with the metrical accent, are probably the most distinctive features of *candombe* rhythm [158].

5.3 Micro–rhythmic analysis

This section is devoted to measuring and analysing the micro–rhythmical properties of the drumming patterns in *candombe*. Unlike the other two drums, the *chico* drum is characterized by a single pattern that must be repeated during the whole performance, establishing the lowest level of the metrical structure. The *repique* drum, on the other hand, is the drum allowed the highest level of improvisation. Its primary pattern (*repique básico*, see Fig. 2.13) may be varied and ornamented in many ways, and alternated with other *repique* or *clave* patterns [110, 158]. It has been noted that in actual performance, the primary pattern presents a per-

[3] In [106], Ferreira includes an ornamented variant of pattern *a* that softly articulates subdivisions 2.2 and 3.2. Apart from that, in [158], Jure introduces another version of the rumba pattern that has an additional stroke at subdivision 3.1.

ceptible deviation with respect to the four pulses of the beat, towards a triplet feeling [157]. And although the *chico* drum is presented as the foundation upon which the whole metrical structure is built, it has been suggested recently that its pattern presents a contraction of the inter–onset intervals [112]. The aim of the methods and experiments proposed in this section is to assess the exact nature of such deviations.

5.3.1 Analysis methods and dataset

Dataset

The multi–track recordings of the dataset described in Section 3.4 were used for the reported experiments—in particular, the 12 ensemble recordings corresponding to groups of three and four drums, i.e. 9 and 3 performances respectively. The total time of the performances was over 35 minutes and the tempos varied between 100 and 140 bpm, with a strong prevalence of values around 130 bpm. Annotations of the location of beats and downbeats for all recordings are available as part of the dataset for beat/downbeat tracking described in Section 3.3.

Onset detection

Automatic detection of onsets was carried out on the separate audio tracks using the methods described in Section 4.3. The resulting events were manually checked and adjusted when necessary, yielding a total of 37062 onsets. A window length of 20 ms and a hop size of 5 ms were used for the STFT computation. This provided a resolution of the detection function that proved to be appropriate for the micro–rhythmic analysis. Some tests were conducted comparing the automatic detected events to the location of their physical onset, which was manually determined by inspecting the audio waveform. For instance, in a *piano* drum track with 396 onsets, less than 5% of them showed an absolute difference to the physical onset greater than 5 ms, and 80% of the differences were below 3 ms. Even at a high tempo value of 140 bpm, a resolution of 5 ms represents less than 5% of the inter–onset interval per metric subdivision, which is approximately 107 ms.

Timing data extraction and analysis

Because of the different tempo of each recording, as well as the variations of the beat pace within a given performance, the timing data can not be analysed as absolute duration values in seconds or milliseconds. Provided that the beginning of each rhythm cycle is identified in some way, a reasonable option is to consider the four-beat rhythm cycle as the reference time interval in order to normalize the timing data. Besides, the location of each onset can be expressed as a percentage of the four-beat rhythm cycle, therefore making it comparable across different recordings and distinct sections of a certain performance.

Then, each onset can be grouped to the closest subdivision within the rhythm cycle. The grid of subdivisions is assumed to be isochronous at this stage, i.e. an equidistant partitioning of the rhythm cycle in four beats, each beat in four subdivisions. Yet, once the onsets are grouped, the mean and the standard deviation of each group can be computed. This provides an estimate of the actual location of the events of the rhythmic pattern for the different metric positions, as well as their amount of dispersion.

It remains to be determined how to find the location of the beginning of each rhythm cycle or downbeat. A straightforward solution is to have the downbeats of the recordings manually annotated by an expert, for instance by tapping to the piece, as is the case for the dataset under study. This, however, may introduce an external subjective bias that could be troublesome when attempting to precisely measure micro–rhythmic deviations.

An alternative method, used in some music entrainment studies [240], is to estimate the location of the downbeat from the onsets themselves. It is based on identifying an ostinato or timeline pattern that articulates the downbeat throughout the whole piece. In this way, a preliminary estimate of the beginning of each rhythm cycle is obtained. After that, the onsets articulated by all the instruments within a tight window around the beginning of each rhythm cycle are considered, and their locations are averaged to provide an estimate of the downbeat, which is hopefully not biased to any particular instrument. Since the normalization is accomplished by using the same time reference for all instruments, it allows the comparison of the location of the onsets among different ensemble parts.

Difficulties arise, however, when an attempt is made to implement this method for *candombe* drumming, since choosing the reference for the downbeat is far from trivial and can be rather problematic. Ideally, it should be the timeline proper, i.e. the *clave* or *madera* pattern, but *candombe* drumming is atypical in that the timeline pattern is not always present. Then, there is the *piano* drum which has a timeline-like function and its *base* patterns articulate the downbeat. Nevertheless, it also exhibits several variations and *repicado* patterns which do not assure an articulation at the downbeat for every rhythm cycle. Finally, the *chico* drum is the tempo/pulse reference for the other instruments, but its most common pattern does not articulate the downbeat (or any other beat, see Fig. 2.12). If micro–temporal deviations of the *chico* drum pattern do actually exist, trying to interpolate the downbeat from the other strokes might be prone to error.

Yet, the preliminary estimate of the beginning of each rhythm cycle can be obtained from the downbeat annotations. The problem arises when trying to estimate the downbeat from the average of the onsets at the beginning of the rhythm cycle, since it may be the case that none of the drums articulate an onsets there. But there is an alternative *chico* drum pattern that adds a soft stick stroke at each beat, thus articulating all subdivisions. Therefore, only the performances in which the *chico* drum plays the alternative pattern are considered for testing the method, since it guarantees at least one onset at the downbeat. Fortunately, there are 6 recordings of this kind in the dataset, exhibiting the alternative *chico* pattern throughout the whole performance. The remaining 6 recordings, in which the *chico* plays the standard pattern, are analysed using only the manual annotations as time reference, and the results produced by both methods are compared.

5.3.2 Experiments and results

Two different types of experiments were carried out to study the micro–rhythmic properties of *candombe* drumming. The first one relies on the manual annotations for normalizing the timing data. In the second one, the location of the downbeat is determined by averaging the onsets lying at the beginning of the rhythm cycle.

Experiment 5.6 The focus of this experiment is to study the micro–rhythmic characteristics of the standard *chico* pattern and the *repique* primary pattern (as depicted in Fig. 2.13), i.e. the group of 'headless' sixteenth notes of the former and the sixteenth–eighth–sixteenth note motif of the latter. Six takes were selected from the dataset, in which the standard *chico* pattern is played. A few introductory rhythm cycles at the beginning of the performances were discarded, that included the *clave* pattern played by all the drums and occasionally some alternative *chico* patterns. The selected takes feature the three performers that played both *chico* and *repique* during the recording session. The automatic onset detection provided a total of approximately 5000 *chico* onsets for the analysis. In the *repique* tracks, only the segments with the primary *repique* pattern were analysed, resulting in approximately 1500 onsets. The patterns under analysis repeat themselves in each of the four beats of the rhythmic cycle. Hence, the normalization of the timing data was done using the manually annotated beats. Then, the onsets were assigned to its closest subdivision, considering a perfect division in four of the beat. Finally, mean and standard deviation of the onsets were computed for each group. The analyses show some clear tendencies appearing consistently in all cases, although the exact amounts of deviation are different for each take, even of the same player.

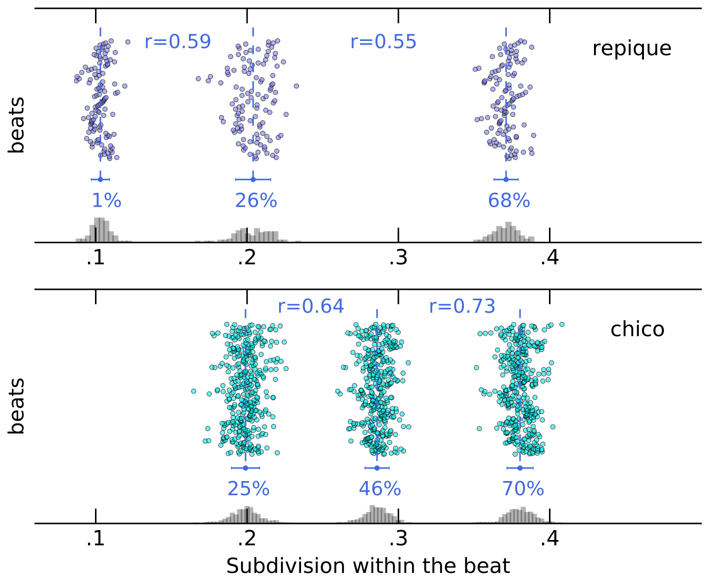


Figure 5.9: Analysis of *chico* and *repique* tracks of the same recording, beats plotted upwards. Mean values of the onsets at each subdivision are given as a percentage of the beat duration. The Pearson correlation coefficient ($p < 0.001$) is provided for pairs of consecutive onsets.

In Fig. 5.9 the results of the analysis of one of the recordings are depicted, showing the typical behaviours of *chico* and *repique* patterns. The last onset of the *repique* primary pattern is displaced as expected, and is very close to a ternary division of the beat. The second onset keeps its place around the second pulse, but has a higher variance. And despite the prevalent descriptions of the *chico* drum as strictly establishing the pulse of the lowest metrical level, its pattern presents a significant temporal contraction: the first onset (the hand stroke) coincides quite precisely with the second pulse, but the two remaining onsets appear clearly ahead of the divisions in four of the beat. Small quantitative differences aside, the same behaviour of both drums was observed in all the analysed recordings.

Experiment 5.7 This experiment examines the micro–rhythm properties of all the ensemble parts of a complete recording. All normalizations of the timing data are done only at the four-beat cycle level, not at each beat independently as in the previous experiment. Six takes of the dataset are selected for this experiment, in which the *chico* pattern softly articulates the beats, thus providing an onset at the downbeat. The total number of detected onsets in the recordings is 19054, of which 8428 correspond to the *chico*, 5773 to the *repique* and 4853 to the *piano*. An estimate of the downbeats, obtained by averaging all the onsets at the beginning of each rhythm cycle, is used as time reference for the normalization. The steps used for processing the timing data are detailed in the following.

1. The beginning of each rhythm cycle is obtained from the annotations
2. All onsets are normalized to relative positions between adjacent cycle starts
3. The onsets are assigned to its closest subdivisions within an isochronous grid
4. Onsets of all drums at first subdivision are averaged to estimate downbeats
5. All onsets are normalized using the downbeat estimates as reference
6. The onsets are assigned to its closest subdivisions within an isochronous grid

Once the onsets are assigned to subdivisions, their mean and standard deviation values are computed. It is worth noting that, as expected, the standard deviation values obtained with the downbeat estimates are smaller compared to that of the preliminary alignment based on annotations.[4]

The results of this process are provided in Fig. 5.10 for one of the recordings of the dataset.[5] Despite the highly structured form in which the onsets are organized, some of them may lie in ambiguous locations with regards to the metric grid. These onsets usually arise from embellishments which do not conform to the stable

[4]It was in order to check this that all the onsets are assigned to subdivisions based on the preliminary alignment, instead of only identifying those corresponding to the downbeat.

[5]The *piano* part in this recording is played with virtuosity, filling the rhythmic patterns with soft strokes in between the accented ones. Although other *piano* drum renditions are typically less saturated, this recording is selected because it better reveals the underlying subdivision of the pulse intended by the performer and allows for a richer discussion.

pattens of interest. For this reason, they are filtered out, by discarding all the events outside a window of three standard deviations to each side of the mean. In Fig. 5.10, the filtered events are marked with crosses. The mean and standard deviation values for each group are computed again after the filtering process. Anyway, the small number of such events that actually come out of the process (less than 1.5%) indicate that they have an almost negligible effect in the analysis.

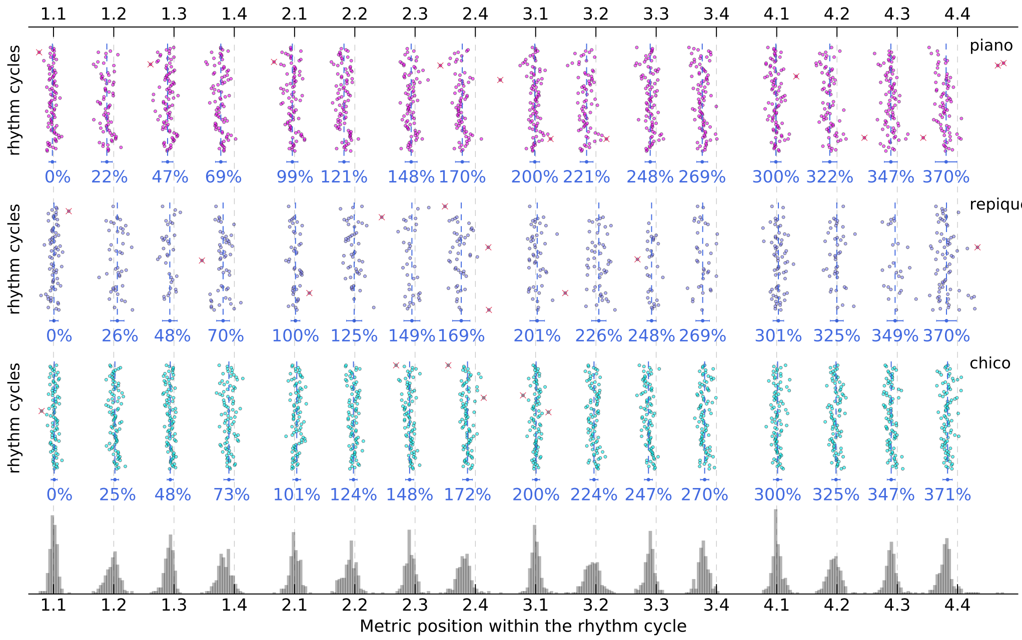


Figure 5.10: Analysis of all the ensemble parts of a complete recording. The onsets are cycle-length normalized using the estimated downbeats. The rhythm cycles are plotted upwards in increasing order. Mean values of the onsets are given as a percentage of the beat duration.

A number of observations that are worth noting emerge from the analysis of this recording. Firstly, considering each of the three drums in turn, it can be noticed that all the beats show a similar behaviour, with differences in the mean values of its four subdivisions of about only one percent of the beat duration. This suggests that in terms of the micro–timing all the beats are alike for the different ensemble parts. Secondly, comparing the timing of the different drums, they considerably adhere to a similar pattern. The onsets at the beats line up quite precisely at an equidistant metrical grid. Conversely, the onsets at the third and fourth subdivisions noticeably depart from a perfect partition in four of the beat, the deviations being pretty consistent across the different ensemble parts. The only remarkable difference is that of the second pulse of each beat. While the *chico* and *repique* drums adhere quite strictly to an equidistant subdivision, the onsets of the *piano* drum lie a little bit ahead. Overall, these results are qualitatively in agreement with those of the other recordings analysed, and those reported in the first experiment for the *chico* and *repique* drums.

The bottom of Fig. 5.10 shows the aggregated histogram of all onsets of the different drums in the recording. Disregarding the small discrepancies between the different ensemble parts, the peaks of the aggregated histogram could be considered as the corresponding metrical grid positions of the piece.

For the purpose of extending this analysis to the whole dataset, an aggregated histogram is computed including all the onsets of the six takes selected for this experiment, which is shown in Fig. 5.11. Not surprisingly, the micro–timing pattern of each beat is remarkably similar. For this reason, the timing data is mapped to the beat length, assuming an equidistant metrical grid of beats within the rhythm cycle, and the aggregated histogram is computed again. The results are presented in Fig. 5.12, along with the aggregated histogram of the *chico* onsets for the six takes of the first experiment for comparison. It can be seen that both experiments, although not strictly comparable, yield very similar micro–timing patterns.

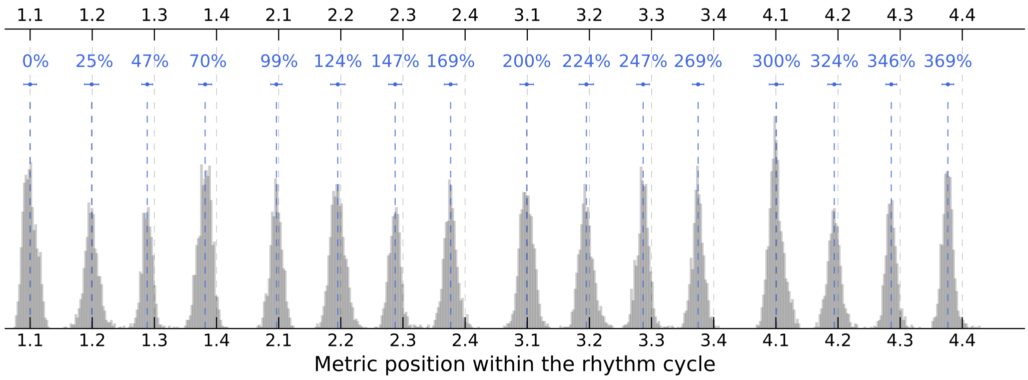


Figure 5.11: Histogram aggregating the onsets of all drums of the six takes selected for Exp. 5.7. Mean values of the onsets are given as a percentage of the beat duration.

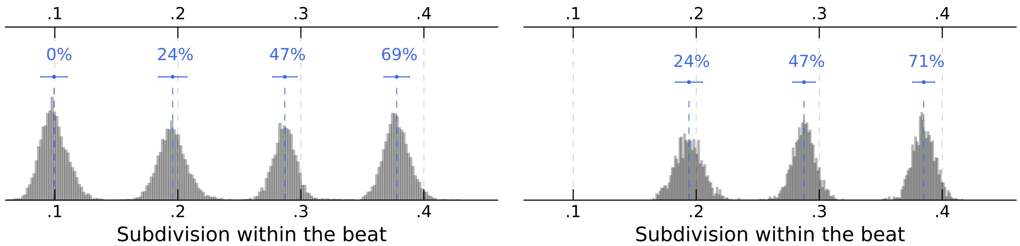


Figure 5.12: Left: aggregated histogram of the onsets of all drums of the six takes of Exp. 5.7. Right: aggregated histogram of the *chico* onsets of the six takes of Exp. 5.6.

These findings, while preliminary, suggest that *candombe* rhythm has an isochronous grid of beats, that exhibits an uneven subdivision structure. Within the beat, the micro-timing follows a short–short–short–long (SSSL) pattern,[6] given by duration proportions of approximately 24:23:23:30; though the exact amounts of deviation depends on the dataset considered and the analysis method applied.

[6] Actually, very close to a medium–short–short–long (MSSL) pattern.

5.4 Discussion and conclusions

In this chapter some methods were proposed for the detailed analysis of the rhythmic patterns of each type of drum found in recorded *candombe* performances.

An interactive software tool was developed which loads an audio file and a set of beat/downbeat labels, and produces in return a bar-length map of rhythmic patterns and a clustering analysis. The usefulness of the proposal was illustrated through a set of experiments concerning the study of *piano* drum performances from audio field recordings. For a given recording, a map of bar-length patterns permits the inspection of their evolution over time and the visualisation of important structural aspects of the performance. Besides, a clustering analysis of the rhythmic patterns detected in the recording tends to match characteristic patterns of the instrument. In addition, a comparison of *piano* drum performance styles was conducted by considering the patterns of the largest cluster (i.e. *base* patterns). Results of the experiments indicate that by applying the proposed methods, patterns tend to be grouped by artist disclosing their personal styles. Moreover, their similarities reveal common traits of the traditional styles and even family ties.

The methods were also applied to the study of the *clave* (or *madera*) patterns from a dataset of audio recordings. Most of the patterns reported in the literature arose from the analysis, which also allows for a quantitative assessment of the number of rhythm cycles of each type and their location within the performance.

In spite of the promising results obtained, the characterization of the rhythm patterns should be further investigated. The classification of the type of stroke for each articulated pulse is envisioned as an important improvement of the technique.

A study of the subdivision timing of the rhythmic patterns was accomplished by using multi-track recordings of *candombe* drumming performances by renowned players. The analysis of several recordings revealed the systematic use of micro–rhythmical deviations in the patterns of the *candombe* drums, indicating that micro–timing is a structural component of its rhythm. The behaviour of the *repique* primary pattern was more precisely measured, and a behaviour of the *chico* pattern that does not fit current descriptions was confirmed [112]. The consistent use of these deviations can be considered as an evidence of the existence of a sort of "swing" characteristic of *candombe*, analogous to the idea of swing in Jazz and other Afro–American music styles [38, 63, 112, 136, 149, 216]. The micro-rhythmic structure of *candombe* rhythm was characterised as an isochronous grid of beats, that exhibit an uneven subdivision following a short-short-short-long (SSSL) pattern.[7] Nonetheless, further experiments should be carried out on a wider dataset including more performers representing the different traditional styles of *candombe*.

[7]This micro-timing profile typologically resembles the subdivision structures in some styles of Afro-Brazilian music, such as maracatu and candomblé [136]. Furthermore, comparing several Brazilian styles, the long fourth "upbeat" subdivision (...L) stands out as an invariant feature [136]. Hence, without denying the differences between the various rhythms, it may be argued that the long fourth subdivision provides some common aesthetic ground across many Afro-Brazilian styles, which potentially includes Afro-Urugayan *candombe*. This was suggested by Rainer Polak in personal communication, April 3th 2017.

Chapter 6

Beat and downbeat tracking

This chapter is based on work originally reported in [219]. The description herein reproduces some passages of the article and includes modifications and additions in order to put the work in the context of this dissertation. Besides, a different observation model introduced into the proposed method improves the reported results, and some other state–of–the–art beat/downbeat tracking algorithms that became available after the publication of the paper are added to the comparison.

6.1 Introduction

Meter plays an essential role in our perceptual organization of music. In modern music theory, metrical structure is described as a regular pattern of points in time, hierarchically organized in metrical levels of alternating strong and weak beats [181, 192]. The *beats* specifically refer to the pulsation of the perceptually most salient metrical level, which are then further grouped into *measures* or *bars*. The first beat of each measure is called the *downbeat*. The metrical structure itself is not present in the audio signal, but is rather inferred by the listener through a complex cognitive process. Therefore, a computational system for metrical analysis from audio signals must, explicitly or implicitly, make important cognitive assumptions. A current cognitive model proposes that, given a temporal distribution of events, a competent listener infers the appropriate metrical structure by applying two sets of rules: Metrical Well-Formedness Rules (MWFR), which define the set of possible metrical structures, and Metrical Preference Rules (MPR), which model the criteria by which the listener chooses the most stable metrical structure for a given temporal distribution of events [181]. While not strictly universal, most of the MWFR apply for a variety of metric musics of different cultures [280]; MPR, on the other hand, are more subjective and, above all, style-specific. Hence, a listener not familiar with a certain type of music may not be able to decode it properly, if, for instance, its conventions differ substantially from usual tonal metrical structures.

This is why the computational analysis of rhythmic/metrical structure of music from audio signals remains a difficult task. Most of the proposed algorithms follow a bottom-up approach with little prior information about the music under analysis [93, 99, 166, 220], often including some kind of preference rules—e.g. by aligning beats with onsets of stronger and/or longer events [181]. Although they perform reasonably well for a large part of the popular music of Western origin (such as rock or pop with a steady beat), they usually fail on processing syncopated or polyrhythmic music, for instance, that of certain Turkish, Indian or African traditions [275]. Actually, the identification of challenging music styles and the development of style–specific algorithms for meter analysis and beat–tracking has been regarded as a promising direction of research to overcome the limitations of existing techniques based on a supposedly universal model [77, 232, 267, 275].

Consequently, other approaches follow a top–down process guided by high–level information, such as style–specific characteristics [152, 295]. Among them, the explicit modelling of rhythmic patterns has recently been proposed as a way to improve upon existing beat–tracking algorithms by Krebs et al. [174]. This is based on the Bayesian approach referred to as *dynamic bar pointer model*, first proposed by Whiteley et al. [293], and later extended by various authors [172, 174–176, 276, 294]. In general, the model aims at the joint estimation of beats, downbeats, tempo, meter, and rhythmic patterns, by expressing them as hidden variables in a Hidden Markov Model (HMM) [244]. The joint inference of all the rhythm parameters allows to exploit the mutual dependencies between them, but also increases the computational complexity of the model. For this reason, an efficient transition model was devised in [175] to reduce the dimension of the hidden–variables state-space of the HMM. Besides, the use of particle filtering was also proposed as an efficient inference scheme [176, 276]. In [174], Krebs et al. effectively applied the rhythmic pattern modelling to a dataset of Ballroom dance music, and argued that, since the rhythmic patterns are learned directly from data, the model could be adapted to music of any kind. Then, the observation model was adapted in [141] to a more diverse collection of music from different cultures (Makam music from Turkey, Cretan music from Greece, and Carnatic music from the South of India). Recently, the dynamic Bayesian networks for beat and meter tracking have been combined with Recurrent Neural Networks (RNN) [51,173] and Convolutional Neural Networks (CNN) [96, 143].

The work presented in this chapter was motivated by the fact that some characteristics of *candombe* rhythm are challenging for most of the existing rhythm analysis algorithms. Indeed, when this research work started, we confirmed that none of the publicly available beat tracking methods were able to deal with *candombe* recordings properly. For this reason, in collaboration with Leonardo Nunes—who was doing research on the Bayesian approach for rhythm analysis [218]—the author of this thesis work tackled the development of a beat/downbeat tracking algorithm suitable for *candombe* drumming. Also departing from [293], the method was developed almost in parallel to that proposed by Krebs et al. in [174]. Preliminary results were presented at the CICTeM Congress [253] in 2013, and at the International Musical Rhythm Workshop in 2014 [250]. Finally, the work gave rise to a paper presented at the ISMIR Conference in 2015 [219]. With the publication of this paper, an annotated dataset of *candombe* recordings was released, together with the first—to the best of our knowledge—publicly available software implementation of the Bayesian approach for rhythm analysis.[1] Other two implementations of the Bayesian approach were made publicly available shortly after, namely *madmom* [49] and *bayesbeat* [175], and are included in the experiments in this chapter. However, only the latter allows for training the models with annotated recordings, and thus it was possible to adapt it to *candombe* rhythm.

In the following, the proposed supervised scheme for rhythmic pattern tracking is described, which aims at finding the metric structure from an audio signal, including the phase of beats and downbeats. Then, in Section 6.3, the performance of the proposed method is assessed over the dataset of annotated recordings introduced in Section 3.3, considering different experimental set-ups, and comparing it to that of other available beat and downbeat tracking algorithms. The chapter ends with discussion and conclusions of the presented work.

6.2 Rhythmic pattern matching

A Bayesian paradigm is adopted in this work in which the rhythmic/meter structure is explicitly modelled as a latent state inference problem, as in [293]. Given a sequence of observed data $\mathbf{y}_{1:K}$ the goal is to identify the most probable hidden state trajectory $\mathbf{x}_{0:K}$. An observation model that relates the observations to the hidden variables has to be defined. Besides, a prior distribution over $\mathbf{x}_{0:K}$ have to be postulated. In this framework the posterior distribution over hidden variables is given by the Bayes's theorem:

$$p(\mathbf{x}_{0:K}|\mathbf{y}_{1:K}) = \frac{p(\mathbf{y}_{1:K}|\mathbf{x}_{0:K})\, p(\mathbf{x}_{0:K})}{p(\mathbf{y}_{1:K})}. \tag{6.1}$$

For on-line or potentially real-time applications, the inference can be implemented in a causal form by performing *filtering*, in which given observations up to the present time, distributions of the form $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ are computed. For off-line inference *smoothing* can be performed, which takes into account future as well as present and past observations, $p(\mathbf{x}_k|\mathbf{y}_{1:K})$. This is a retrospective improvement of estimates, so smoothing distributions can be obtained in terms of the corresponding filtering distributions [293].

In this section, a rhythmic/metric analysis algorithm that matches a given rhythmic accentuation pattern to an audio signal is described. It tries to find the time of occurrence of each *tatum* knowing its expected accentuation inside the pattern, thus being able to track not only the beat but also other metrical information. Initially, a tempo estimation algorithm is employed to obtain the beat period (tempo), assumed to be approximately stable throughout the signal. Then, the main algorithm is used to find the phase of the accentuation pattern within the observed signal. Two different observation models are considered, which differ in the way the likelihood of an accentuation pattern is defined.

[1] The software implemented is available from `github.com/lonnes/RhythmicAnalysis`, and can be applied not only to *candombe* recordings but to other music styles, by informing a rhythmic pattern to track as an input parameter.

6.2.1 Audio feature extraction

For audio feature extraction, the approach based on the spectral flux described in Section 4.2 is applied, as summarized in the following with the respective analysis parameter values. First, the Short-Time Fourier Transform of the signal is computed for sequential frames of 20 ms duration, weighted by a Hann window, in hops of 10 ms. Then, the STFT is mapped to the Mel–scale using 160 bands. No logarithm magnitude compression is applied. The resulting sequences are differentiated (via first-order difference) and half-wave rectified.

For tempo estimation, the feature values are summed along all Mel sub-bands, in order to take into account events from any frequency range.

Since its pattern is the most informative on both *tactus* beat and downbeat locations, the rhythmic pattern tracking is tailored towards the *piano* (i.e. the low–pitched) drum. Therefore, the accentuation feature used for pattern matching is obtained by summing the spectral flux along the lowest Mel sub-bands (up to approximately 200 Hz) only. This function is normalized by the 8-norm of a vector containing its values along ± 4 estimated *tatum* periods around the current time frame (i.e. a window length of half a bar). Recall that the resulting feature value is expected to be close to one if a pulse has been articulated and close to zero otherwise. In addition, it also carries some information on the type of articulation. For instance, an accented stroke produces a higher feature value compared to a muffled one, since in the former case the spectral change is more abrupt.

6.2.2 Tempo Estimation

For tempo estimation, this work adopts a straightforward procedure based on locating the maximum of a suitably defined similarity function. As proposed in [233], the basic function is the product between the auto-correlation function and the Discrete Fourier Transform of the features computed for the whole signal. The result is weighted by the function described in [202]. The period associated with the largest value in this weighted similarity function is selected as the tempo of the signal. After the tempo is obtained, the *tatum* period used for pattern tracking can be computed just by dividing the beat period by 4. This *tatum* period is then used to define the variables in the pattern tracking algorithm as described in the next sections.

6.2.3 Variables definition

In order to perform its task, the algorithm employs two discrete random variables. The first one, called *tatum counter*, $\mathbf{c}_k$, counts how many frames have passed since the last *tatum* has been observed at frame k. Assuming an estimated *tatum* period of τ frames, then $\mathbf{c}_k \in \{0, 1, \dots, \tau - 1 + \sigma_c\}$, where σ_c is a parameter that allows for possible timing inaccuracies in the *tatum*. The second, called *pattern index*, $\mathbf{a}_k$, indicates the position inside a given rhythmic pattern at frame k in the range

$\{0, 1, \ldots, M-1\}$, where M is the length of the rhythmic pattern in *tatums*.[2] The rhythmic pattern will be expected to define a series of accents or lacks of accent in the *tatums*. Time evolution of these two variables will be described in the next section, where it is assumed that the sampling rate of the feature (typically less than 100 Hz) is much lower than that of the original signal (usually 44.1 kHz). The model describes the accentuation feature extracted at frame k as a sample or observation y_k of random variable $\mathbf{y}_k$.

A summary of the proposed model for rhythmic pattern tracking is depicted in Fig. 6.1, where the statistical dependencies among the variables—which are described in the following two sections—are explicitly shown.

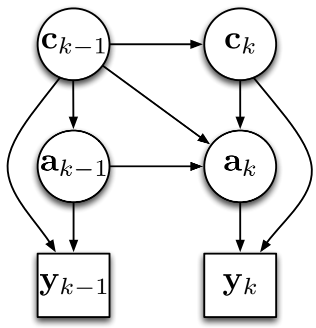


Figure 6.1: Graphical representation of the statistical dependency between random variables and observations. Rectangles denote continuous variables and circles discrete variables.

6.2.4 State Transition

In this section, the probabilities of each value for the two random variables at frame k given past frames are described. A first-order Markov model will be assumed for the joint distribution of the random variables, i.e., the probability of each possible value of a random variable at frame k depends only on the values assumed by the variables at the previous frame, $k-1$. Using this assumption, the two random variables will constitute a Hidden Markov Model [214].

The *tatum* counter variable, as previously mentioned, counts how many frames have passed since the last *tatum*. The state $\mathbf{c}_k = 0$ is considered the "tatum state" and indicates that a *tatum* has occurred at frame k. This random variable is closely related to the *phase state* proposed in [91] for beat tracking. Only two possible transitions from frame $k-1$ to frame k are allowed: a transition to the "tatum state" or an increment in the variable. The transition to the "tatum state" depends on both the past value of the variable and the (known) *tatum* period. The closer the value of the variable is to the *tatum* period, the more probable is the

[2]Since bar–length rhythmic patterns are used in the reported experiments, it also indicates the position within the bar at frame k.

transition to the "tatum state." Mathematically, it is possible to write

$$p_{\mathbf{c}_k}(c_k|c_{k-1}) = \begin{cases} h[c_{k-1} - (\tau - 1)], & \text{if } c_k = 0 \\ 1 - h[c_{k-1} - (\tau - 1)], & \text{if } c_k = c_{k-1} + 1 \\ 0, & \text{otherwise,} \end{cases} \tag{6.2}$$

where $h[.]$ is a tapering window with $h[n] = 0$ for $|n| > \sigma_c$ that models possible timing inaccuracies on the *tatum*, and $\sum_n h[n] = 1$. Currently, a normalized Hann window is employed to penalize farther values. The value $\sigma_c = 2$ was set for the reported experiments, indicating that inaccuracies of up to about 50 ms are tolerated by the algorithm.[3]

Since the accentuation pattern is defined in terms of the *tatum*, its time evolution will be conditioned by the pattern evolution. Assuming that the pattern indicates the expected accentuation of the next *tatum*, the variable should only change value when a "tatum state" has been observed, indicating that a different accentuation should be employed by the observation model (described in the next section). Hence, mathematically

$$p_{\mathbf{a}_k}(a_k|c_{k-1}, a_{k-1}) = \begin{cases} 1, & \text{if } (a_k = a_{k-1} \oplus 1) \wedge (c_{k-1} = 0) \\ 1, & \text{if } (a_k = a_{k-1}) \wedge (c_{k-1} \neq 0) \\ 0, & \text{otherwise,} \end{cases} \tag{6.3}$$

where $\wedge$ is the logical AND, $\oplus$ denotes a modulo-M summation, and M is the length of the accentuation pattern. As can be gathered, given the previous *tatum* counter value, the pattern index becomes deterministic, with its next value completely determined by its value at the previous frame and the value of the *tatum* counter. The transitions for this variable are inspired on the ones used in the family of algorithms based on [293] (i.e. [48,141,174]), except for defining the pattern in terms of *tatums* instead of an arbitrary unit.

The joint distribution of the two variables at frame k can be obtained as

$$p_{\mathbf{c}_k,\mathbf{a}_k}(c_k, a_k|c_{k-1}, a_{k-1}) = p_{\mathbf{a}_k}(a_k|c_{k-1}, a_{k-1}) p_{\mathbf{c}_k}(c_k|\mathbf{c}_{k-1}). \tag{6.4}$$

6.2.5 Observation Model

This section describes the likelihood of $\mathbf{c}_k$ and $\mathbf{a}_k$ given an observed accentuation $\mathbf{y}_k$ in the signal. We consider two different options for defining the likelihood function. The first idea, presented in [219], is to measure the difference between the expected accentuation (provided by the rhythmic pattern to track) and the observed one. The larger the difference, the less probable the observation. Then, the second option is to directly estimate the likelihood function from the labeled data using kernel density estimation techniques.

[3] This tolerance is consistent with the micro–timing deviations reported in Section 5.3, whose maximum values spanned from approximately 20 to 30 ms depending on tempo.

Likelihood based on expected accentuation pattern

If the accentuation pattern is a vector $\overline{A} \in \mathrm{R}^{M \times 1}$ containing the expected feature values, then at frame k the likelihood for $c_k = 0$ ("tatum state") can be defined as

$$p_{\mathbf{y}_k}(y_k|c_k, a_k) = N_{\sigma_t}(y_k - \overline{A}_{a_k}), \tag{6.5}$$

where $N_{\sigma_t}(.)$ is a Gaussian function with zero mean and variance σ_t^2 used to model possible deviations between expected and observed accents. For $c_k \neq 0$, the likelihood is given by:

$$p_{\mathbf{y}_k}(y_k|c_k, a_k) = N_{\sigma_d}(y_k), \tag{6.6}$$

where N_{σ_d} is a zero-mean Gaussian with variance equal to σ_d^2. Hence, the closer to zero the feature, the more probable the observation. This is similar to the non-beat model adopted in [91], and is not found in [174, 293].

In the reported experiments, $\sigma_t = \sigma_d = 0.5$, thus allowing for a reasonable overlap between expected and actual observed values.

Estimation of the likelihood function

The dataset of annotated recordings can be used to directly estimate the likelihood of the observed accentuation variable $\mathbf{y}_k$, given the values of the state variables $\mathbf{c}_k$ and $\mathbf{a}_k$. If only the *tatum* state is taken into account, i.e. $\mathbf{c}_k = 0$, the histogram of the low-frequency feature for each of the sixteen possible values of the pattern index $\mathbf{a}_k$ is depicted in Fig. 6.2.[4]

A possible approach for the modelling would be to fit a theoretical distribution to the data, such as a gamma distribution. However, in some cases there are more than one local maxima and there seems not to be a single theoretical distribution appropriate for all the *tatums*. Another option would be to model the observation probabilities with a mixture of Gaussian distributions (GMM) as in [141, 174]. But, given the skewness of the data and the fact that the support is bounded, a Gaussian distribution is not able to accurately fit the boundaries. Therefore, the Kernel Density Estimation (KDE) approach is preferred in this work, which is a standard non–parametric technique to estimate the probability density function of a random variable without making rigid assumptions about the distribution of the data [139]. Even though the bounded support can also hinder the KDE approach, some techniques have been devised to overcome the boundary bias issue, as described in the following.

Let $\{X_1, X_2, \ldots, X_n\}$ be a sample of independent and identically distributed observations from an unknown distribution F_X with associated density f_X; the standard kernel density estimator is

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{6.7}$$

[4]Note that the distribution of values is quite consistent with the simplified basic form of the *piano* drum introduced in Section 2.3.2. For instance, the most skewed distribution towards 1 corresponds to the accented stroke of the fourth *tatum*. Besides, the histograms also show that the third and fourth *tactus* pulses are usually articulated.

where K is the *kernel* function, a non-negative function that integrates to one and has zero mean and usually is symmetric and unimodal, and h is the *bandwidth* parameter that controls the smoothness of the estimation [139]. The selection of the appropriate value for h can be carried out by means of cross–validation. A variable bandwidth can also be devised, such that different values for h are selected to locally adapt to different portions of the data [270].

When the support of f_X is bounded, the estimator places some positive mass outside the support for x close to the boundary, preventing the estimation to be consistent in those areas [118]. For this reason, some techniques have been proposed to properly handle the bounded support I of f_X, either by using a kernel function also supported on I, as in the *Beta kernel estimator* [72], or by transforming the variable of interest into another whose density estimation should be free from boundary problems and then transforming the results back to the original bounded support, such as in the *probit–transformation* for kernel estimators [118].

After some preliminary experiments in which standard KDE, variable bandwidth, Beta kernel and probit–transformation estimators were tested, the latter was selected as the best performing option. In Section 6.3.7 experiments are reported on the performance achieved by the proposed algorithm with the likelihood function estimated by standard KDE and probit-transformation methods.

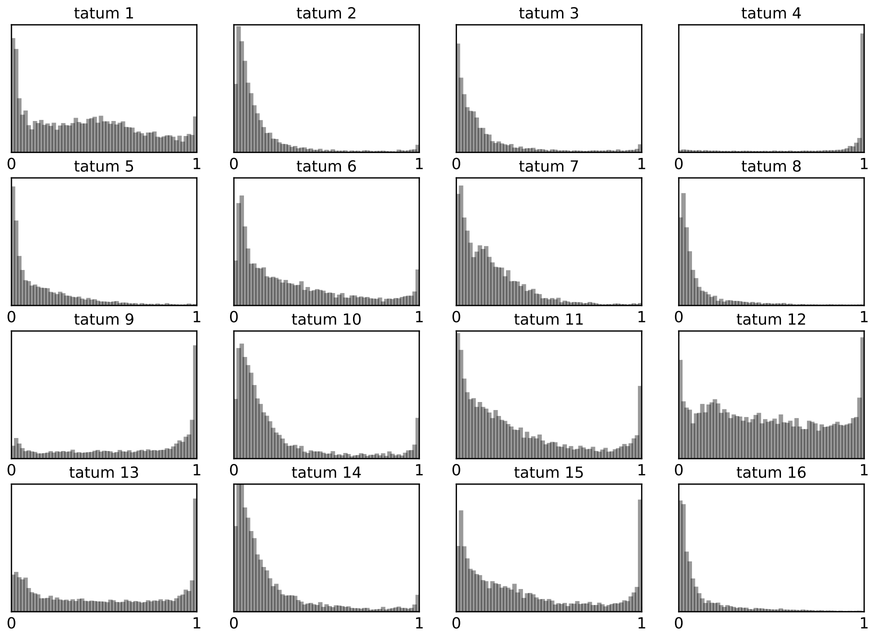


Figure 6.2: Histogram of feature values at each *tatum* computed for the whole dataset.

6.2.6 Inference method

Different inference strategies can be employed to find the most probable pattern index and *tatum* counter values given the observed accentuation, such as the forward–backward algorithm [214, 293]. In this work, the well-known Viterbi algorithm [214] is employed to find the most probable path among all possible combinations of values of each random variable given the observed features y_k.

6.2.7 Prior state distribution

At last, a uniform prior is chosen for $\mathbf{c}_0$ and $\mathbf{a}_0$ indicating that the counter and the pattern can start with any possible value in the first frame.

6.3 Experiments and results

A set of experiments was devised to assess the performance of the proposed rhythmic pattern tracking system with respect to the problems of estimating the rate and phase of beats and downbeats, using a dataset of manually annotated *candombe* recordings. Several state-of-the-art beat-tracking algorithms were included in the experiments in order to evaluate how challenging *candombe* drumming is for approaches that do not exploit a priori information about the rhythm [48, 51, 93, 99, 166, 220, 296]. Besides, the implementation of the Bayesian approach described in [174] was trained and tested with the annotated dataset for comparison.

6.3.1 Dataset

The dataset of annotated *candombe* recordings introduced in Section 3.3 was used for the reported experiments. Whenever a learning phase was part of the algorithm, such as in the estimation of the likelihood function or the selection of a rhythmic pattern to inform from the data, a leave–one–out scheme was used [139].

6.3.2 Tempo estimation

Since tempo estimation is only an initialization step of the rhythmic pattern tracking task, whose overall performance will be examined in detail, it suffices to mention that the estimated tempo was within the interval spanned by the annotated beat periods along each of the files in the dataset, thus providing a suitable value for the respective variable. The estimated tempo value for each file and the corresponding median value of the annotated beats are compared in Fig. 6.3. Higher differences actually correspond to recordings with non-stable tempo, whose impact on the performance of the proposed algorithm is discussed in Section 6.3.7.

6.3.3 Performance measures

Among the several objective evaluation measures available for audio beat tracking [81], there is currently no consensus over which to use, and multiple accuracies

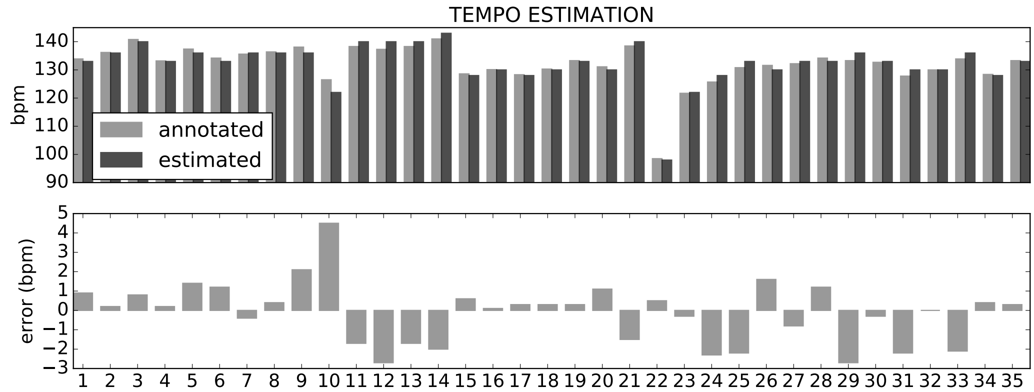


Figure 6.3: Comparison of the median value of the annotated beats and the estimated tempo.

are usually reported [48, 80]. In a recent pilot study, the highest correlation between human judgements of beat tracking performance and objective accuracy scores was attained for CMLt and Information Gain [80].

In this work CMLt, AMLt and F-measure were adopted, as their properties are well understood and were considered the most suitable for the current experiments. The non-inclusion of Information Gain was based on the observation that it yielded high score values for estimated beat sequences that were definitely not valid. Specifically, in several instances when the beat rate (or a multiple of it) was precisely estimated, even if the beat phase was repeatedly misidentified, the Information Gain attained high values while other measures such as CMLt or F-measure were coherently small. In the following, a brief description of the adopted metrics[5] is provided (see [81] for details), along with the selected parameter values.

The CMLt measure (Correct Metrical Level, continuity not required) considers a beat correctly estimated if its time-difference to the annotated beat is below a small threshold, and if the same holds for the previous estimated beat. Besides, the inter-beat-interval is required to be close enough to the inter-annotation-interval using another threshold. The total number of correctly detected beats is then divided by the number of annotated beats and expressed as a percentage (0-100 %). Both thresholds are usually set to 17.5 % of the inter-annotated-interval, which was also the value adopted in this work. The AMLt measure (Allowed Metrical Levels, continuity not required) is the same as CMLt but does not take into account errors in the metrical level and phase errors of half the period.

The F-measure (Fmea) is the harmonic mean of precision and recall of correctly detected beats, where precision stands for the ratio between correctly detected beats and the total number of estimated beats, while recall denotes the ratio between correctly detected beats and the total number of annotated beats. A beat is considered correctly detected if its time-difference to the annotation is within ± 70 ms; this tolerance was kept in this work.

Only CMLt and F-measure were used for assessing the downbeat, since the

[5]Computed with standard settings using code at `https://code.soundsoftware.ac.uk/projects/beat-evaluation/`.

loosening of metrical level and phase constraints in AMLt was considered inappropriate. The parameters used for the downbeat evaluation were the same as those used for the beat, i.e. 17.5 % of the inter-annotated-interval and ±70 ms tolerance.

6.3.4 Experiments with general–purpose beat tracking algorithms

Some rhythmic characteristics of *candombe* drumming are potentially troublesome for the typical algorithmic approaches for beat and downbeat tracking. As noted in Section 2.3.2, the *chico* drum pattern defines the pulse, but it usually does not articulate the *tatum* that falls on the beat and has instead a strong accent on the second. Besides, the *clave* divides the 16-*tatum* cycle irregularly, with only two of its five strokes coinciding with the beat. Finally, the strong phenomenological accents displaced with respect to the metric structure add to the difficulty, such as the accented *piano* stroke at the fourth *tatum* of the first beat.

Fig. 6.4 shows two examples of the typical behaviour of general–purpose beat–tracking algorithms when dealing with the *candombe* recordings of the dataset. The strong accent of the *chico* drum is sometimes predicted as the location of the *tactus* beat, as can be seen in the first example, in which the predictions are approximately aligned around the second *tatum* beat. The characteristic strong accent performed by the piano drum at the fourth *tatum* beat is also problematic. In the second example, this accented stroke is the first strong audio event at the beginning of the plot. Then, it can also be identified in the next measure, four tactus beats ahead, and once again almost at the end of the plot. As can be seen, some algorithms tend to predict the beat at this point. In addition, there are also some errors in detecting the correct beat period.

All the beat–tracking algorithms listed in the following were considered for the evaluation experiments: the algorithm based on dynamic programming[6] by D. Ellis [99], two systems based on multi-agents, namely BeatRoot[7] by S. Dixon [93] and INESC-BT[8] by J.L. Oliveira et al. [220], the multi–feature beat tracker by J.R. Zapata et al. [296] provided by the Essentia library,[9] two algorithms by S. Böck et al., both based on recurrent neural networks (RNN) and dynamic Bayesian networks (DBN), namely DBNBeatTracker [48] and DBNDownBeatTracker [51], as provided by the madmom library [49],[10][11] and the meter analysis algorithm based on a bank of comb filter resonators and a hidden Markov model by A. Klapuri et al. [166], which was kindly provided by the author.

Table 6.1 shows the performance attained by each of the algorithms, and also (for conciseness) the experiments discussed in the next sections. Results are averaged over the whole database and weighted by the number of beats and downbeats

[6] `http://labrosa.ee.columbia.edu/projects/beattrack/`

[7] `https://code.soundsoftware.ac.uk/projects/beatroot`

[8] `http://smc.inesctec.pt/technologies/ibt/`

[9] `http://essentia.upf.edu`

[10] `https://github.com/CPJKU/madmom`

[11] Note that the implementation of [48] provided by madmom does not use the multi–model based on RNN, i.e. it corresponds only to the DBN version of the algorithm. The efficient state space and transition model for the DBN described in [175] are used.

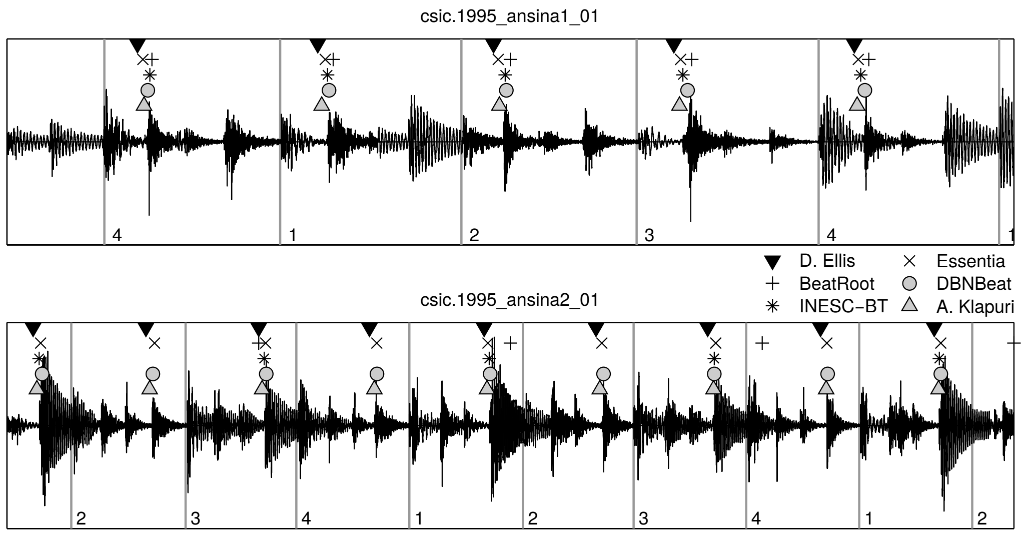


Figure 6.4: Two examples of the typical difficulties encountered by beat-tracking algorithms: D. Ellis [99], BeatRoot [93], INESC-BT [220], Essentia [296], DBNBeat [48] and A. Klapuri [166]. The vertical lines show the annotated beats, numbered within a four-beat measure, while their predicted location is depicted with different markers for each algorithm.

of each audio file. Note that two of the algorithms are able to provide estimates for the downbeat [51, 166]. Although the beat rate (or a multiple) is sometimes precisely estimated by the general–purpose beat–tracking algorithms, the correct metrical level and/or the phase of the beat is usually misidentified.

6.3.5 Experiments with informed rhythmic patterns

Two different experiments are conducted to evaluate the proposed system with the first observation model, i.e. the likelihood is based on an expected accentuation pattern $\overline{A}$ that is informed as an input parameter. This section describes the first type of experiment, in which the pattern to track $\overline{A}$ is informed to the algorithm based on musical knowledge about the rhythm, without any training or tuning to data. On one hand, this has a practical motivation: even when no annotated data is available one could take advantage of the technique. On the other hand, it gives a framework in which musical models can be empirically tested. In short, an informed rhythmic pattern based on musical knowledge is nothing but a theoretical abstraction, and this type of experiment could provide some evidence of its validity.

To that end, based on the different ways the *piano* pattern is notated by musicology experts [106, 158], a straightforward approach was adopted. Firstly, the *piano* pattern as introduced in Fig. 2.13 (usually regarded as the *piano* in its minimal form) was considered. A binary pattern $\overline{A}$ was assembled by setting a value of 1 for those *tatums* which are expected to be articulated and 0 otherwise. Then, a more complex pattern was considered by adding two of the most relevant articulated *tatums* which were missing, namely the 6th and 15th, and also building the corresponding binary pattern. Hence, the binary informed patterns are

$$\text{Pattern 1:} \quad \overline{A} = [1,0,0,1,0,0,0,0,1,0,0,1,1,0,0,0]$$
$$\text{Pattern 2:} \quad \overline{A} = [1,0,0,1,0,1,0,0,1,0,0,1,1,0,1,0].$$

Considering binary patterns is certainly an oversimplification of the real rhythmic patterns, since it does not take into account the accented and muffled strokes that are an essential trait of a *piano* performance. It would be possible to encompass dynamic variations into the informed pattern by considering distinct quantized values of the feature for different type of strokes. However, the binary patterns were favoured for the sake of simplicity and as a proof of concept. Next section deals with rhythmic patterns that are not binary.

	BEAT			DOWNBEAT	
	CMLt	AMLt	Fmea	CMLt	Fmea
General–purpose – Section 6.3.4					
D. Ellis [99]	44.2	63.0	43.8	–	–
BeatRoot [93]	13.9	14.9	22.7	–	–
INESC-BT [220]	9.1	27.6	16.7	–	–
Multi-feature Essentia [296]	39.0	39.6	37.6	–	–
DBNBeatTracker [48]	14.3	18.8	13.9	–	–
DBNDownBeatTracker [51]	6.6	11.9	5.7	17.7	0.6
A. Klapuri [166]	28.8	35.5	29.3	36.6	13.2
Informed patterns – Section 6.3.5					
Pattern 1	80.2	80.5	81.3	84.7	79.1
Pattern 2	79.0	81.0	79.8	81.2	77.5
Learned patterns – Section 6.3.6 (leave-one-out)					
Median	79.9	79.9	80.8	82.4	76.9
K-means 2	81.7	81.7	82.6	84.4	79.3
K-means 5	82.5	82.5	83.6	85.2	80.6
Density estimation – Section 6.3.7 (leave-one-out)					
Standard KDE	89.2	89.2	91.4	90.5	89.2
Probit-transformation	91.4	91.4	93.4	92.6	91.6
Bar pointer model – Section 6.3.8 (leave-one-out)					
Bayesbeat (B=1, R=1, T=2006) [174, 293]	98.6	98.6	99.1	100.0	100.0
Bayesbeat (B=1, R=1, T=2015) [174, 175]	98.9	98.9	99.3	100.0	100.0
Bayesbeat (B=1, R=2, T=2015) [174, 175]	98.9	98.9	99.3	100.0	100.0
Bayesbeat (B=2, R=1, T=2015) [174, 175]	99.0	99.0	99.3	100.0	100.0

Table 6.1: Performance of all the different experiments conducted. Results are averaged over the whole dataset, weighted by the number of beats and downbeats of each audio file.

6.3.6 Experiments with learned rhythmic patterns

In this section, the second type of experiment with the first observation model is described, in which the accentuation pattern $\overline{A}$ is based on rhythmic patterns actually present in real performances, extracted from the annotated dataset. There are different possible approaches to extract a single rhythmic pattern to track from the annotated data. Firstly, for each *tatum*-grid position in the bar-length pattern, all the feature values in the dataset are collected. The distribution of feature values in the low-frequency range will be dominated by the *base* patterns of the *piano* drum, albeit there will be a considerable amount of *repicado* patterns [254]. In order to cope with that, $\overline{A}$ is assembled using the median of feature values for each *tatum* beat, which is less influenced by outliers than the mean.

The problem with the median pattern is that it models different beat positions independently. A better suited approach is to group the patterns based on their similarity into a given number of clusters,[12] and select the centroid of the majority cluster as a good prototype of the *base* pattern. This was described in Section 5.2 and applied in [254], to identify *base* patterns of the *piano* drum. Fig. 6.5 shows the patterns learned from the whole database, using the median and the centroid of the majority cluster obtained with K-means for 2 and 5 clusters. It is remarkable that the differently learned patterns are quite similar, exhibiting the syncopated 4th *tatum* beat as the most accented one. The locations of articulated beats for the informed patterns of the previous section are also depicted, and are quite consistent with the learned ones. The K-means approach turned out to be little sensitive to the number of clusters, yielding similar patterns for K from 1 to 6.

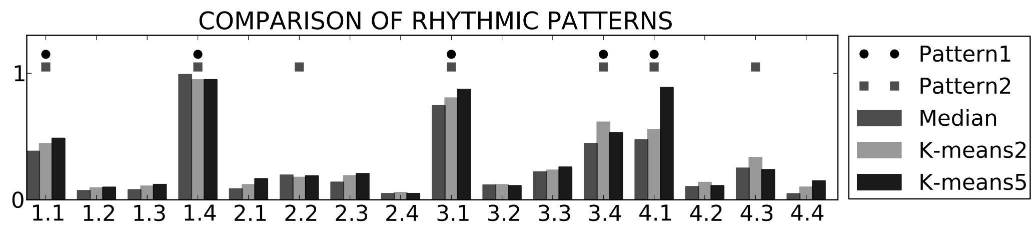


Figure 6.5: Comparison of the different rhythmic patterns considered for the experiments of sections 6.3.5 and 6.3.6. Median and K-means patterns are learned from the whole dataset.

The performance of the approach was assessed using a leave-one-out scheme and the results are detailed in Table 6.1. Not surprisingly, performance is almost the same for the different rhythmic patterns. Considering different feature values instead of binary patterns did not yield any notable performance increase.

A detailed inspection of the performance attained for each recording in the database, depicted in Fig. 6.6, shows there is still room for improvement, given that about half-a-dozen files are definitely mistracked. This may indicate that the pattern $\overline{A}$ to track simply does not properly match the given performance. To check this hypothesis, a K-means (K=2) clustering was carried out only with the

[12] Grouping rhythmic patterns into clusters from annotated data is used in [141,174], as detailed in [172], to adapt the dynamic bar pointer model to specific music styles.

candidate patterns found within each target recording, whose tracking was then performed using the centroid of the majority cluster as $\overline{A}$. In Fig. 6.6, the new results obtained for the files with lower performance (CMLt<50%) in the dataset are depicted in colour. Except for the first one, performance was (sometimes notably) improved when the informed rhythmic pattern is the one that better matches the recording. Therefore, modelling several patterns as in [141, 174] can potentially improve the current results, even though other source of error exist.

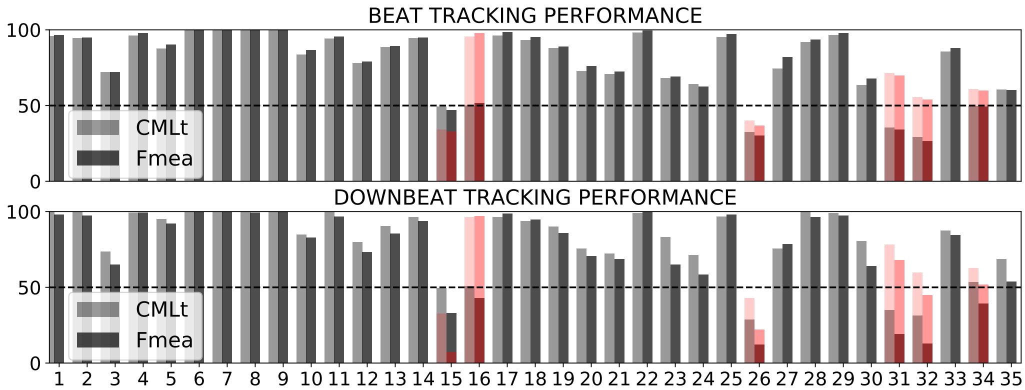


Figure 6.6: Leave-one-out performance for each file using K-means pattern with K=2. Results with the centroid of the majority cluster for the low performing files are depicted in colour.

6.3.7 Experiments with likelihood density estimation

This section is devoted to describing the assessment of the proposed algorithm when using the second observation model, in which the likelihood function is estimated from the annotated data by using a KDE technique. The likelihood of the "tatum state", i.e. $\mathbf{c}_k = 0$, for the different values of the pattern index $\mathbf{a}_k$ (as shown in Fig. 6.2), and the likelihood of the "non-tatum state", i.e. $\mathbf{c}_k = 1$, have to be estimated. Two different methods are considered, the standard KDE and the probit–transformation technique for dealing with the $[0, 1]$ bounded support [118].

For each *tatum*, the bandwidth value of the standard KDE method was selected independently, using 3-fold cross-validation, within a range of values from 0.001 to 0.025 with 0.001 step. The linear kernel was the best performing among the different kernel functions tested (linear, Gaussian, exponential), so it was selected for the experiments with the standard KDE. An improved probit–transformation kernel density estimator based on local likelihood estimation was used in the experiments, implemented by locally fitting a 2-degree polynomial to the log–density, which is equivalent to a local Gaussian estimation [118]. A local adaptive bandwidth in the transformed variable is achieved by a k-nearest-neighbour method [118].

The results obtained with the KDE approach are provided in Table 6.1. The most important remark is that the performance is notably increased by the adoption of the observation model based on the estimation of the likelihood function. Compared to the previous model, based on informing an accentuation pattern $\overline{A}$

as input parameter, the improvement is about 10% in all performance measures considered. This indicates that the system takes advantage of the more flexible observation model in order to deal with the variability of the rhythmic patterns present in real recordings. Besides, the use of an estimation technique that can properly handle the bounded support of the feature variable provides a small but consistent increase in all the performance measures.

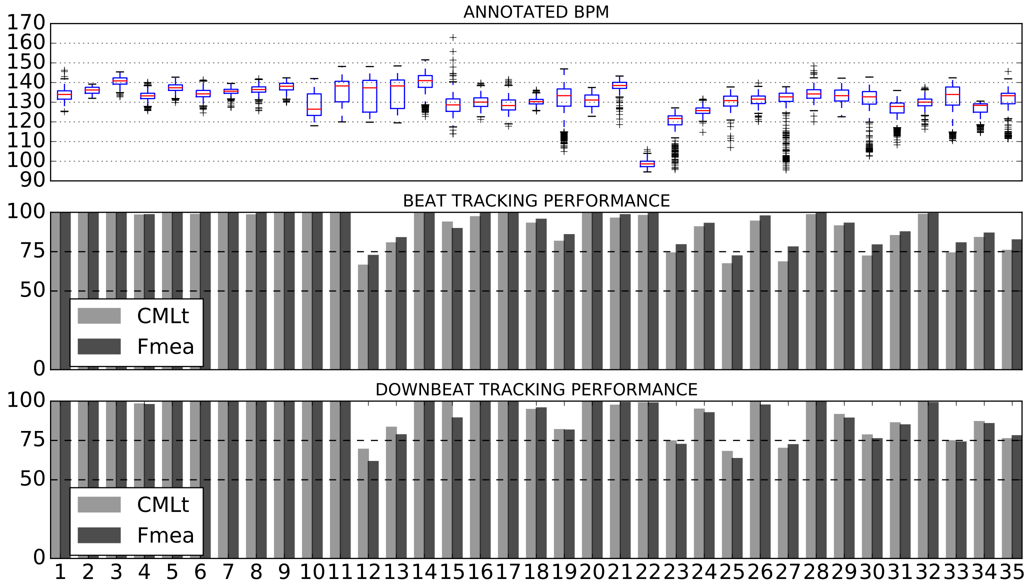


Figure 6.7: Leave-one-out performance for each file using probit–transformation estimation and their corresponding box-plots of annotated bpms.

Notwithstanding the increase of the overall results, the performance of some files is somewhat less that 75%, as shown in Fig. 6.7. The inspection of these low performing files reveals that most of them show tempo changes that are difficult to handle by the proposed algorithm, since it it assumes that tempo is approximately stable through the piece. The definition of the state variables, in particular the *tatum counter* $\mathbf{c}_k$ is based on an initial tempo estimate. Hence, despite that certain tolerance for timing inaccuracies in the *tatum* is taken into account (see Section 6.2.3), this restricts the capability of the algorithm to properly deal with tempo changes. Fig. 6.7 also depicts the box-plots of the tempo values extracted from the manual annotations. It can be seen that there exist certain correlation between the variability of the tempo values and the performance attained by the algorithm. For instance, the first 9 files of the dataset exhibit a quite tight distribution of tempo values while achieving high performance measures. Conversely, the tempo curves of two of the lowest performing files (number 12 and 27) cover a wide range of values, as illustrated in Fig. 6.8 along with the estimated tempo. Their behaviour clearly depart from the hypothesis of an approximately stable tempo, and are therefore troublesome for the proposed algorithm. Regardless of the observation model utilized by the algorithm, the ability to properly deal with tempo variations stands as the most necessary improvement to turn the system able to increase current performance. An implementation of the dynamic bar pointer model that handles tempo variations is considered in the next section.

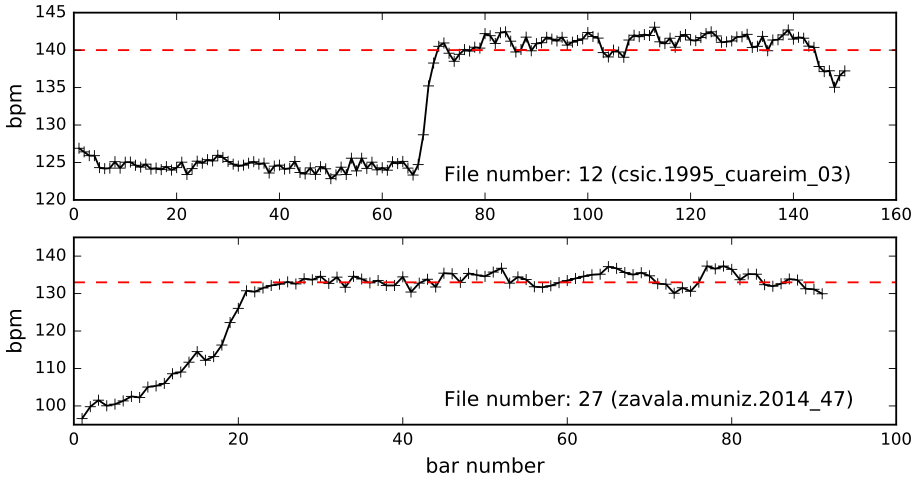


Figure 6.8: Tempo curves for two low performing files, estimated value shown with a line.

6.3.8 Experiments with trained dynamic bar pointer model

This section describes the experiments conducted with *bayesbeat*[13], a package for metrical structure analysis of musical audio signals based on the dynamic bar pointer model [293], that includes all the extensions described in [141, 172, 174–176, 276]. Among several other features, it allows for training a model from an annotated audio dataset, and therefore it was adapted in this work to track beats and downbeats in *candombe* recordings. Next, a brief comparison to the proposed algorithm is provided that highlights the most relevant differences. After that, the experiments conducted and the results obtained are discussed.

Compared to the proposed algorithm, the model implemented in *bayesbeat* that was used in the experiments [174], has several differences that are worth noting. First and foremost, the state space model includes—apart from a variable m_k that tracks the position inside a bar (similar to $\mathbf{a}_k$, defined in Section 6.2.3)—a discrete variable n_k that models the tempo value, and an index r_k that selects one among the rhythmic patterns present in the dataset [174]. For this reason, the model extracts the tempo value at each frame k and is able to track its variations, within a given range and according to a certain transition model detailed in [174]. Besides, several rhythmic patterns are modelled, though a rhythmic pattern is assumed to remain constant throughout the whole piece [174].[14]

Note that there is no variable such as the *tatum counter* $\mathbf{c}_k$ of the proposed algorithm, that connects the temporal grid of the audio frames to the subdivisions

[13] `https://github.com/flokadillo/bayesbeat`

[14] The transition model was extended in [141] to allow transitions between rhythmic pattern states within a piece, whenever the bar pointer m_k crosses a bar border.

in the metric structure, similar to the beat phase state in [91]. In consequence, the models differ in the way the rhythmic patterns are defined within the bar: in terms of *tatum* pulses for the proposed model (see Section 6.2.3), or using a 64-bins grid to divide the measure (for a 4/4 meter) in [174]. Although the observation features used in [174] are also based on the spectral flux, as in the proposed algorithm (see Section 6.2.1),[15] the observation model has some differences. The most substantial one is the use of a two-dimensional feature, comprising low-frequencies (<250 Hz) and high-frequencies (>250 Hz), instead of considering only a low-frequency band as in the proposed algorithm. Besides, the likelihood function is obtained by fitting a GMM to the feature values of each bin in the one-bar grid [174].

Various test were conducted to study the influence of the different parameters, some of which are described in the following. A leave–one–out scheme was used for training an testing. Instead of learning the tempo range from the data, a fixed tempo range yielded better results, using the default interval of 60 to 230 bpm. The two different transition models available were tested, the original one from [293] (T=2006) and the efficient model proposed in [175] (T=2015). According to the results in [141], using one rhythmic pattern per rhythm class is usually enough to achieve a good performance and provided the best results in most cases.[16] This was also the case for *candombe*, the use of one or two patterns (R=1, R=2) actually yielded the same results. Therefore, the use of a single rhythmic pattern was adopted for most of the experiments. The importance of using two-dimensional features was investigated by considering only the low frequencies (B=1) and both low- and high-frequency bands (B=2). The default value of I=2 number of components was used for the GMM, and the inference was done with the Viterbi method.

The results obtained are presented in Table 6.1. The performance attained is virtually perfect according to the measures considered, and represents a notable improvement with respect to the proposed algorithm. Using only the low-frequency band (B=1) and a single rhythmic pattern (R=1), as in the proposed algorithm, was certainly sufficient to accurately track beats and downbeats in *candombe* recordings. This indicates that, not surprisingly, the ability to track tempo changes is a critical issue for improving the performance of the proposed algorithm. Besides, the use of the efficient transition model has no negative impact on the results, as also noted in [175]. The finer grid in which the rhythmic patterns are represented and the fact that the "tatum state" model is not needed, may have also a positive effect on the performance and deserve a thorough study.

Fig. 6.9 shows the rhythmic pattern learned by the model in the low-frequency band from the whole dataset. It depicts the normalised mean of the feature values for each of the positions within a bar in the 64th-bin grid. The distribution of feature values is quite consistent with that of the patterns used in the previous experiments (see Fig. 6.5). For instance, the articulated strokes of Pattern 1 from Section 6.3.5, exactly match the five most salient positions of the rhythmic

[15]The feature extraction is actually very similar, except for the use of logarithmic magnitude compression (which emphasises higher frequencies and has no relevant effect if only low frequencies are considered) and the normalization process.

[16]Note that updated results and errata were made available after the publication.

pattern of Fig. 6.9. It is also worth noting that the distribution of feature values follows a similar micro–timing pattern to that reported in Section 5.3. While the articulated strokes at the beats show their maximum feature value closely aligned to their corresponding subdivision (i.e. 1.1, 3.1 and 4.1), the accented syncopated stroke at the 4th *tatum* (i.e. 1.4) exhibits its maximum value in the bin just before, indicating that it is ahead of the exact subdivision in four of the beat (and this also holds for the stroke at 3.4).

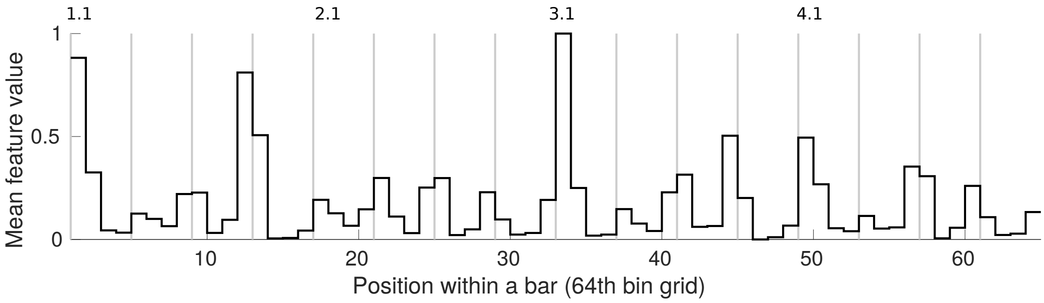


Figure 6.9: Rhythmic pattern learned from the whole dataset in the low-frequency band.

6.4 Discussion and conclusions

This chapter tackled the problem of automatic rhythmic analysis of *candombe* drumming from audio signals. From the rhythm description and the presented experiments, it becomes clear that typical assumptions of general–purpose beat–tracking algorithms (such as strong events at beat times) do not hold, which hinders their performance. In order to overcome this problem, an algorithm based on the Bayesian approach for rhythmic analysis [293] was proposed. By tracking a rhythmic pattern in the low–frequency band that informs when a beat with or without accentuation is expected to occur, the algorithm can eventually determine the complete metric structure. Indeed, experiments employing both rhythmic patterns based on musical knowledge and others learned from the annotated dataset, showed that the proposed algorithm can estimate the beat and downbeat positions correctly for most of the files of the dataset, attaining an overall CMLt score of about 80 to 90 % depending on the observation model applied.

In its present form, the proposed algorithm has a limited ability to properly deal with tempo changes, which probably constitutes its main drawback. Certain correlation has been shown between the low performance attained for some recordings and the variability of their tempo curve. Actually, a more elaborate algorithm [174], also based on the Bayesian approach, that is able to track tempo variations showed a virtually perfect performance when trained and tested with the dataset of *candombe* recordings. These results provide further evidence for the need of properly dealing with tempo changes in order to increase the current performance of the proposed algorithm. Besides, the experiments also reinforce the idea that modelling a single rhythmic pattern in the low–frequency band, which corresponds to the *piano* drum, is sufficient to allow the inference of the whole metric structure in *candombe* drumming.

In general, therefore, the present work gives additional evidence of the generalizability of the Bayesian approach to complex rhythms from different music traditions. Moreover, the obtained results are very encouraging and allow us to confidently tackle other problems and applications that rely on automatic tracking the metric structure from *candombe* recordings.

Chapter 7

Analysis based on information theory

7.1 Introduction

Often a parallel is drawn between data compression and computational learning [163, 187]. The argument runs as follows: the more we are able to compress the data, the more we have learnt about their underlying regularities. Data compression works by discarding irrelevant information and exploiting repeating patterns. This is essentially the ability to generalize from specific examples to general rules, which is arguably the very definition of learning. The derivation of general rules from specific examples is known as *inductive inference* and has deserved a great deal of theoretical and experimental work [19, 274]. In turn, a learning algorithm can be judged by how succinctly it explains the observed data. This is a form of *Occam's razor* principle, which states that, assuming that all other things are equal, a shorter explanation for the observed data should be preferred over a longer one.[1] A formalization of this paradigm in the form of a model of algorithmic learning is called Occam learning [163].

Similarly, it is also common to draw a connection between data compression and complexity assessment [180, 297], albeit the actual meaning of complexity may be domain specific and difficult to seize without a formal definition. Simply put, data compression captures the amount of structured information present in a certain phenomenon, therefore the compression ratio can serve as a measure of the complexity of the data. This idea has been applied in a myriad of disciplines, such as bioacoustics [164], linguistics [155], image processing [248], and biomedical signal analysis [2], among others. Actually, complexity has been granted central roles in psychological models of aesthetic appreciation [42], where the greater the degree of uncertainty an artistic stimuli contains, the greater the amount of information it conveys. This form of subjective complexity assessment is often applied to

[1]This is by no means an irrefutable principle of logic, and there are examples where Occam's razor would have favoured the wrong theory given the available data [187]. In science, parsimony is used as heuristic technique for the development of theoretical models. The preference for simplicity is based on the falsifiability criterion, that is, simpler theories are preferable to more complex ones because they are more testable [117].

music [98,198,278], and relies on objective measures that are usually derived from information theory.

Music understanding can be partly thought of as a problem of finding repeated patterns, and thence structure [151,168,215]. Ultimately, data compression can be tailored to the problem of explicitly finding structure through repeated patterns in the data under analysis [147]. There is some recent work in this line for the analysis of symbolic representations of musical pieces by applying different general purpose text compression techniques [196] and point-set compression algorithms [204, 205]. The latter treat each note of a music score as a point in pitch-time space and generate encodings that can be interpreted as detailed thematic/motivic analyses and compared with those produced by musicologists. Based on such encodings, the Normalized Compression Distance (NCD) [186] can be used to measure similarity between pieces for classification purposes, e.g. of folk songs into tune families [204]. Generally speaking, the distance between two pieces is small if one can be significantly compressed given the information in the other. A similar approach is followed in [54], where pattern discovery in a folk song dataset is applied to compression and classification into tune families. The NCD is also used in [73,74] to cluster musical styles and composers from symbolic representations of musical pieces. Likewise, Lempel-Ziv compression [297] is applied in [191] to cluster symbolic musical pieces for studying musical style and authorship. Indeed, the use of Lempel-Ziv compression for music style modelling from symbolic sequences can be traced back to [26]. Lately, the same authors moved to the use of compression methods based on Factor Oracle [14] for learning musical sequences [25].

Consequently, information theory—and particularly, data compression—stands as an appealing framework for music modelling. Source coding is a mapping from a sequence of symbols from an information source to a sequence of alphabet symbols, such that the source symbols can be exactly recovered (*lossless coding*), or recovered with some distortion (*lossy coding*). In his foundational work on information theory in 1948, Claude Shannon established the limits to possible data compression [269]. His source coding theorem states that it is not possible to losslessly compress the data using an average number of bits per symbol (i.e. coding rate) smaller than the *entropy* of the source. In this context, *entropy* is the expected value of the information contained in each message, where information is defined as the cologarithm of the symbol probabilities.

Information theory aims at providing a measure of the amount of information in the data, which can be interpreted as the length of its most compact description. In this approach, the messages or objects to be encoded are supposed to be outcomes of a known random source, whose characteristics determine the encoding. Therefore, given a random source of known characteristics, we are interested in the minimum expected number of bits per symbol to transmit a message from the source through an error-free channel.

This is closely related to the *Kolmogorov complexity* theory [135, 183], also known as *algorithmic information* theory, which was independently introduced by R. J. Solomonoff [274], A. N. Kolmogorov [170], and G. J. Chaitin [70] in the 1960s. Intuitively, the Kolmogorov complexity of a sequence of data is the length of the shortest computer program that can generate it [187]. The computer program is then the compressed version of the data, and the length of such a program can be used as a measure of its complexity. In this case, the amount of information in an object is also related to the length of its description. However, unlike the Shannon information theory, which takes into account the characteristics of the random source of which the object is one of the possible outcomes, only the object itself is considered to determine the number of bits in its compressed version irrespective of how the object arose [135]. Unluckily, the Kolmogorov complexity is not computable [187], i.e. given an arbitrary sequence of data there is no algorithm that returns the shortest program able to produce the data. In practice, computable approximations have to be adopted.

A practical alternative is the *minimum description length* (MDL) principle, introduced by J. Rissanen in 1978 [249], and significantly developed since then in both the mathematics and applications [32, 134]. It is also concerned with the greatest compression of the data by means of a certain model, and considers a lossless representation that also takes into account the cost of describing the chosen model itself. Then, among a collection of different candidate models, the one that achieves the shortest description length of the data is selected. In that sense, MDL can also be regarded as a practical implementation of the Occam's razor principle, and constitutes a powerful method for inductive inference and model selection.

Recently, some research work has been devoted to exploiting the methodology of MDL and Kolmogorov complexity for the analysis of symbolic music. In [201], the author proposes an expression for MDL complexity of Hidden Markov Models (HMMs), that can be applied to symbolic modelling of music structure. In [203] an approach based on Kolmogorov complexity is discussed when applied to the description of symbolic music. Somehow related, in [271] a given piece of music is treated as a sequence of symbols and the shortest possible context-free grammar that generates it is looked for. The use of probabilistic grammars for symbolic music analysis in this framework is reviewed in [1].

For lossy source coding, Shannon introduced and developed the theory of source coding with a fidelity criterion, also called rate-distortion theory, which provides the theoretical foundations for lossy data compression [268,269]. In practice, when we have a continuous source we are not necessarily interested in exact recovery, but only in approximate recovery within a given tolerance. Hence, a distortion measure is introduced to account for the average of the information loss. The problem of coding is then formulated as determining the minimal number of bits per symbol, as measured by the coding rate, so that the source can be approximately recovered without exceeding a given distortion value. Rate-distortion theory has been studied in the information-theory community for more than fifty years [40, 41]. Today, rate-distortion theoretical concepts are an important component of many lossy compression techniques and standards, and have been successfully applied to lossy coding of speech, high-quality audio, images, and video [41,225]. Besides, model selection can also be tackled via rate-distortion theory if the requirement for lossless coding is relaxed [161, 171]. In addition, rate-distortion properties of individual objects can be analysed by means of the recently developed *algorithmic* *rate-distortion theory* [290], which, inspired by the Kolmogorov complexity of individual finite objects, lays down the foundations for an algorithmic analogue of Shannon's probabilistic rate-distortion theory [90]. Nevertheless, the application of a lossy source coding scheme to the description of music and the analysis of its structure, even though suggested in a recent seminar on computational music structure analysis [46], remains, to the best of our knowledge, virtually unexplored.

The main motivation of this work arises from a novel idea: to recast the downbeat detection task as a data compression problem. To do that, different possible alignments of the beats within the rhythm cycle are evaluated and a parsimony criterion is used to select the one corresponding to the downbeat. The hypothesis is that the correct alignment will allow for a simpler explanation of the data compared to the misaligned ones. For this purpose, a lossy compression framework based on the rate-distortion theory is adopted. This is done because, instead of using symbolic music, the analysis is performed on a continuous data source, which corresponds to an accentuation feature function that is directly computed from audio recordings. In this way, a sort of music structure analysis problem—in its minimal expression—is formulated based on the rate-distortion theory. Additionally, it turned out that the obtained description is well suited for addressing other related tasks, namely the complexity assessment of performances and the estimation of the number of different rhythmic patterns in a given recording.

The rest of the chapter is organized as follows. Next section introduces the rate-distortion theory and describes the methods applied. In Section 7.3 the proposed approach to deal with audio recordings of *candombe* performances is presented. Some experiments and the obtained results are reported in Section 7.4. The chapter ends with a critical discussion on the present work, including promising directions for future research.

7.2 Rate-distortion theory

The *rate-distortion* theory is usually introduced by noticing that the description of a real number requires an infinite number of bits, thus a finite representation of a continuous random variable X can never be perfect [78]. Despite the fact that it is not perfect, we can still try to determine how good the representation is, so the definition of some sort of evaluation measure is actually needed. This is accomplished through the introduction of a measure of *distortion*, namely D, which describes the distance between the random variable and its representation. Thus, by allowing some distortion D, the amount of bits used in the representation can be lowered, a formulation that in fact perfectly fits the notion of *lossy compression*. Adopting a communication theory perspective, this is equivalent to the problem of determining the minimal number of bits per symbol, measured by the bit-rate R, that should be transmitted over a channel so that the input signal can be approximately reconstructed at the receiver without exceeding a certain mean distortion D. Thus, a communication system involving an encoder and a decoder can be formulated based on rate-distortion.

Encoding

Consider such a rate-distortion encoder/decoder system applied to a random variable X. Let $X^n = X_1, X_2, \ldots, X_n$ be a sequence i.i.d $\sim p_X(\mathrm{x}),\ \mathrm{x} \in \mathcal{X}$. This source sequence $X^n \in \mathcal{X}^n$ is represented by the encoder as an index $f_n(X^n) \in \{1, 2, 3, \ldots, 2^{nR}\}$. The decoder represents X^n by an estimate $\hat{X}^n \in \hat{\mathcal{X}}^n$. A $(2^{nR}, n)$-rate distortion code can be defined, which consists of an encoding function,

$$f_n : \mathcal{X}^n \to \{1, 2, 3, \ldots, 2^{nR}\}, \tag{7.1}$$

and a decoding or reproduction function,

$$g_n : \{1, 2, 3, \ldots, 2^{nR}\} \to \hat{\mathcal{X}}^n. \tag{7.2}$$

The rate-distortion encoder/decoder system defined so far is presented in Fig. 7.1. The decoded sequence $g_n(f_n(X^n)) = \hat{X}^n$ is a quantized version of the original source sequence X^n. Then, a vector quantization scheme that is optimal with regards to the distortion measure has to be considered.

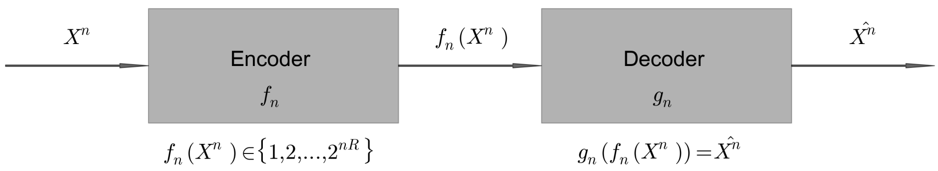


Figure 7.1: Rate-distortion encoder/decoder.

Vector quantization

If we are given R bits per symbol to represent the source X, the problem is to find the optimum reproduction points which minimize the distortion measure, i.e. design a vector quantizer. A vector quantizer Q is a mapping from an Euclidean space of dimension k, $\mathcal{R}^k$, to a finite set $\mathcal{C}_M$, known as *codebook*, which contains M output vectors $\hat{\chi}_m \in \mathcal{R}^k$, known as *codevectors*,

$$Q : \mathcal{R}^k \to \mathcal{C}_M = \{\hat{\chi}_1, \hat{\chi}_2, \ldots, \hat{\chi}_M\}. \tag{7.3}$$

Associated with each codevector $\hat{\chi}_m$ there is a reconstruction region or cell that can be defined as

$$\mathcal{R}_m = \{\mathrm{x} \in \mathcal{R}^k \mid Q(\mathrm{x}) = \hat{\chi}_m\}. \tag{7.4}$$

The encoder is completely specified by the partition of $\mathcal{R}^k$, and the decoder is completely specified by the codebook. Given a distortion measure, the mean distortion value achieved by the system is computed as

$$D = \int_{\mathcal{R}^k} d(\mathrm{x}, Q(\mathrm{x}))\ p_X(\mathrm{x})\ \mathrm{dx}. \tag{7.5}$$

Two simple properties are useful to find an appropriate vector quantizer. First, given a set of reconstruction points $\{\hat{\chi}_i\}$, the distortion is minimized by mapping each element of X^n to the closest of them, which is known as the *nearest-neighbour* condition. The set of regions defined by this mapping is called a *Voronoi* or *Dirichlet* partition, and correspond to the nearest-neighbour regions with respect to the distortion measure. Then, given a certain partition, the selection of the reconstruction point for each region should minimize the distortion measure. This is accomplished by selecting the centroid of the region as the reconstruction point, which is known as the *centroid* condition. The *generalized Lloyd* algorithm for designing a vector quantizer is based on these properties [188, 189].

Generalized Lloyd algorithm

The generalized Lloyd algorithm [188] is an iterative algorithm that starts with a certain set of reconstruction points and finds the optimal reconstruction regions as the nearest-neighbour regions with respect to the distortion measure. Then, the optimal reconstruction points are computed as the centroids of the reconstruction regions and the iteration is repeated for the new reconstruction points. In this way, the expected distortion is decreased in each iteration, and the algorithm converges to a local minimum in a finite number of iterations. A stopping criteria has to be applied, for instance by means of a threshold δ in the amount of distortion decrease between iterations. The algorithm is summarized in Algorithm 1.

Algorithm 1 Generalized Lloyd algorithm

step 1: start with $m = 1$, initial codebook $\mathcal{C}_1 = \{\hat{\chi}_1\}$ and distortion D_1
step 2: given codebook $\mathcal{C}_m$ find codebook $\mathcal{C}_{m+1}$ by
  2.a finding nearest-neighbour regions $\{\mathcal{R}_m\}$ to partition $\mathcal{R}^k$
  2.b setting reconstruction points $\{\hat{\chi}_{m+1}\}$ as centroids of $\{\mathcal{R}_m\}$
step 3: compute distortion measure D_{m+1} for new codebook $\mathcal{C}_{m+1}$
if $(D_m - D_{m+1} > \delta)$ **then**
  goto step 2

It is worth noting the close relationship between the generalized Lloyd algorithm and the K-means clustering algorithm [150]. The latter also repeatedly finds the centroid of each set in the partition, and then re-partitions the input according to which of these centroids is closest. But the main difference is that K-means clustering operates on a discrete set of points instead of a continuous region. Thus, repartitioning the input means simply determining the nearest centroid to the finite set of points, whilst the generalized Lloyd algorithm actually partitions the whole space into regions. However, note that if the input is a finite set of points both algorithms are equivalent.

In practice, some details have to be taken into account if the input is a finite set of points, for instance, to deal with points that are equidistant to more than one centroid (zero probability boundary condition[2]) or to check for an empty region. Besides, there is a strong dependence on the initial codebook, and different initialization strategies can be followed, such as random selection, pruning or splitting [120].

Distortion measure

The distortion function $d(\mathrm{x}, \hat{\mathrm{x}})$ is a measure of the cost of representing symbol x by symbol $\hat{\mathrm{x}}$. It can be regarded as a mapping

$$d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathcal{R}^+ \tag{7.6}$$

from the set of pairs of the source alphabet and the reproduction alphabet into non-negative real numbers. To measure the distortion between sequences x^n and $\hat{\mathrm{x}}^n$, the average of the per symbol distortion of the elements of the sequence can be used, which is computed as

$$d(\mathrm{x}^n, \hat{\mathrm{x}}^n) = \frac{1}{n} \sum_{j=1}^{n} d(\mathrm{x}_j, \hat{\mathrm{x}}_j). \tag{7.7}$$

Some of the most common distortion functions are the *Hamming* distortion and the *squared–error* distortion. Given $\mathrm{x}, \hat{\mathrm{x}} \in \mathcal{R}^k$, such that $\mathrm{x} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_k]$ and $\hat{\mathrm{x}} = [\hat{\mathrm{x}}_1, \hat{\mathrm{x}}_2, \ldots, \hat{\mathrm{x}}_k]$, the squared–error distortion is the squared 2-norm of the difference between symbols,

$$d(\mathrm{x}, \hat{\mathrm{x}}) = \frac{1}{k} \left\| \mathrm{x} - \hat{\mathrm{x}} \right\|_2^2 = \frac{1}{k} \sum_{i=1}^{k} (x_i - \hat{x_i})^2. \tag{7.8}$$

Operational rate-distortion curve

The relationship between rate and distortion can be characterized by the so-called *rate-distortion function*, $R(D)$, that determines the set of possible achievable points in the rate-distortion trade-off for a certain statistical source class [225]. In order to derive such bounds the source has to be properly characterized, but this can be troublesome for complex sources, such as audio and video signals.[3] Besides, the bound provided by a theoretical rate-distortion function gives no constructive procedure for attaining that optimal performance.

Instead, a practical quantization scheme can be examined, and the best operating points of this particular system can be searched for. If all possible quantization choices for that system are considered for a certain source (described by a

[2]This is in fact a necessary condition for a quantizer obtained by means of the generalized Lloyd algorithm to be optimal, together with the nearest-neighbour condition and the centroid condition. In the case of a continuous input the boundary has zero volume and hence the probability is also zero.

[3]In fact, the rate-distortion function is known in closed form only for special cases, such as the Gaussian source with squared-error distortion or the binary memoryless (Bernoulli-p) source with Hamming distortion [78]. For other distributions, usually numerical methods have to be applied [225].

statistical model or a training set), an *operational rate-distortion curve* can be defined [120, 225]. The curve depicts for each rate the distortion achieved by the best encoder-decoder pair designed for that rate. The points in the curve are said to be operational, since they are achievable with the chosen quantization implementation for the available data. The curve allows to identify the best achievable operating points and to differentiate them from those that are sub-optimal or unachievable. When we can make the search among a fixed and discrete set of parameters, each combination of parameters gives a certain R-D pair, producing a curve of individual admissible operating points. In this case, the convex hull of the set of operational points defines the boundary between achievable and non-achievable performances [225]. Fig. 7.2 depicts an operational rate-distortion curve.

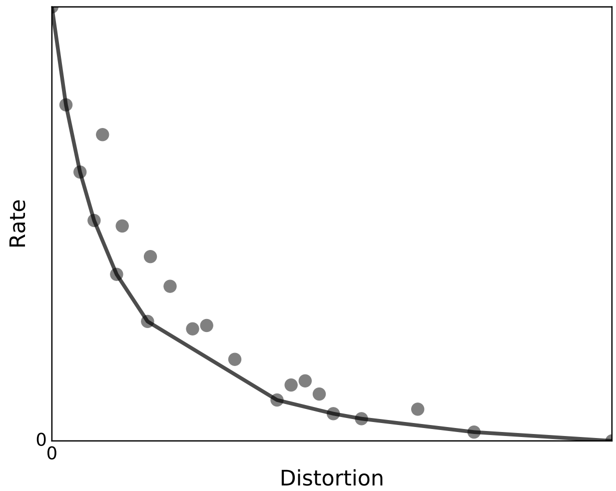


Figure 7.2: Schematic diagram of an operational rate-distortion curve, with the operational points and their convex hull.

Optimization

Within this rate-distortion framework, given a source with a certain distribution and a distortion measure, we seek to establish what is the minimum expected distortion at a particular rate, or equivalently, what is the minimum rate description required to achieve a certain distortion [78]. This can be posed in the form of constrained optimization problems. That means considering either a cost function D with constrained rate $R \leq R_c$, or conversely a cost function R with constrained distortion $D \leq D_c$.

The solution of a constrained optimization problem can often be found by using the so-called Lagrangian method, which minimizes an unconstrained cost function that is the sum of the original objective function and a term that incorporates the constrain and a multiplier $\lambda \geq 0$, a non-negative real number known as Lagrange multiplier. This is a well known technique for problems where the cost function is continuous and differentiable. Yet, when the operational rate-distortion curve is considered, the optimization can be performed through a discrete version of the Lagrangian method [225]. The technique is able to find and optimal solution as long as there exists a point in the convex hull that meets the required constraint. Let l be an index used to denote the operational points on the convex hull of the curve, such that as l increases the rate decreases and the distortion increases. The discrete optimization problem can be formulated as

$$\underset{l}{\text{minimize}} \quad J = D_l + \lambda R_l. \tag{7.9}$$

For a particular value of the multiplier λ, the Lagrangian rate-distortion functional J is minimized as follows. Consider the line contours of constant J value, which are the lines of slope equal to $-1/\lambda$, that are represented graphically in Fig. 7.3 for a certain λ value. The minimization corresponds to finding the point in the convex hull that intersects the line contour corresponding to the smallest J value.

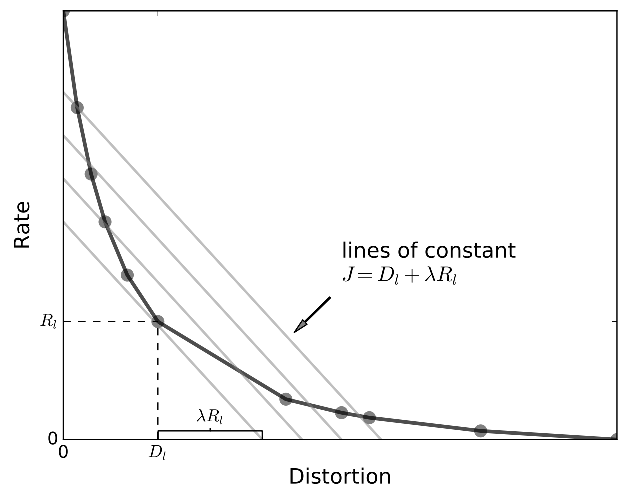


Figure 7.3: Discrete version of Lagrangian optimization for an operational rate-distortion curve.

The Lagrange multiplier λ allows for the selection of specific optimal points in the rate-distortion trade-off. Note that minimizing the Lagrangian J when $\lambda = 0$ is equivalent to minimizing the distortion, whereas minimizing the Lagrangian when λ becomes arbitrarily large is equivalent to minimizing the rate. Intermediate values of λ determine intermediate operating points. Finding the correct value of λ that provides an optimal solution at the required rate can be done using approaches such as the bisection search [225].

7.3 Proposed approach

The proposed approach is based on the idea of describing a complete percussion performance by using a rate-distortion coding scheme. We expect that by studying the coding trade-off between the number of bits per symbol and the amount of distortion we gain some insight into the characteristics of the performance.

An audio recording of a complete *candombe* performance is represented using the spectral audio features described in Section 4.2 and considered as the primary input source to encode. To roughly separate the rhythmic patterns of the different drums, a sub-band filtering is applied. Only the low-frequency band—which corresponds to the *piano* drum—is used in the reported experiments. A local amplitude normalization is carried out to preserve intensity variations of the patterns while discarding long-term fluctuations in dynamics. Assuming the beat positions are known, either manually labelled or automatically tracked, the feature signal is time-quantized by considering a grid of *tatum* pulses.

The resulting input space $\mathcal{R}^k$ has dimension $k = 16$, corresponding to the number of *tatums* in each rhythm cycle. Thus, the input vectors are of the form

$$\mathrm{x} = [x_1, x_2, \ldots, x_{16}]. \tag{7.10}$$

Since features are normalized, each component x_i takes values in $[0, 1]$. A complete performance of length N rhythm cycles is represented by the sequence

$$\mathrm{x}^N = \{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_N\}. \tag{7.11}$$

This sequence can also be regarded as a matrix

$$\mathrm{X} = (\mathrm{x}_{i,j}), \quad i \in [1, 16], \;\; j \in [1, N], \tag{7.12}$$

where i is the *tatum* index and j is the rhythmic cycle index. An example of this type of matrix is presented in Fig. 7.4, for one of the performances from the dataset for beat/downbeat tracking introduced in Section 3.3. Note this representation is nothing but the map of feature patterns proposed in Section 4.2.3.

The vector quantization is implemented following the approach of the generalized Lloyd algorithm. For this particular case, in which the input is a finite set of points, this corresponds to the K-means clustering algorithm. Therefore, each rhythmic pattern x_j of the performance is clustered to a particular group $\mathcal{R}_m$ and represented by its centroid $\hat{\chi}_m$. Fig. 7.4 shows with different colours the grouping obtained for a codebook of size $M = 4$. The input sequence x^N is represented by the encoded sequence $\hat{\mathrm{x}}^N$ only comprising elements of the codebook $\mathcal{C}_M$.

A distortion value, $d(\mathrm{x}_j, \hat{\chi}_m)$, is computed between every pattern symbol x_j of the sequence and its corresponding codevector $\hat{\chi}_m$, using the squared-error distance defined in Equation 7.8. Then, the distortion of the whole input sequence x^N is obtained by averaging the per symbol distortion, using Equation 7.7.

The bit-rate R of the encoded sequence $\hat{\mathrm{x}}^N$ is computed as

$$R = -\sum_{m=1}^{M} p_m \log_2(p_m) \tag{7.13}$$

where M is the codebook size and p_m is an estimate of the probability of occurrence of each symbol. The probability estimate p_m is obtained as

$$p_m = \frac{n_m}{N}, \quad n_m = \#\{\hat{\chi}_m = \hat{\mathrm{x}}_j\} \;\; \forall\, \hat{\mathrm{x}}_j \in \hat{\mathrm{x}}^N, j \in [1, N], \tag{7.14}$$

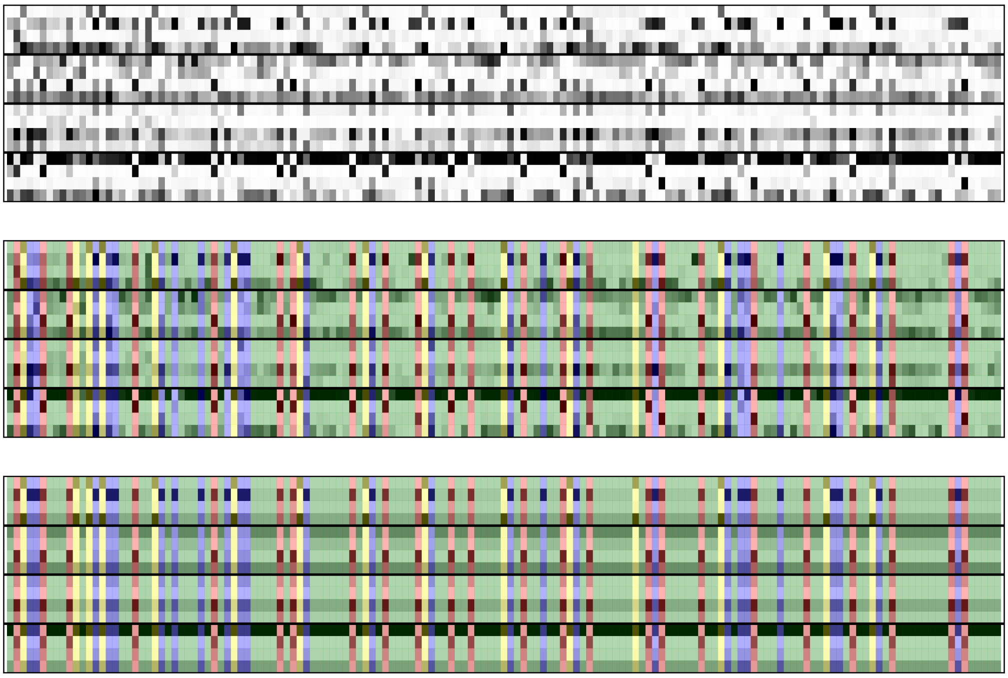


Figure 7.4: Coding process of a complete performance. The input sequence x^N (top), the clusters shown with colours (middle) and the output sequence $\hat{\mathrm{x}}^N$ of codevectors (bottom).

where # denotes the cardinality of the set, and thus n_m represents the number of occurrences of the codevector $\hat{\chi}_m$ in the encoded sequence $\hat{\mathrm{x}}^N$, which is normalized by the total length of the sequence.

The coding process described so far relies on a single parameter, namely the codebook size M. Therefore, an operational rate-distortion curve is obtained by varying the codebook size M and computing the corresponding values for the rate and the distortion. An example of this type of operational curve is depicted in Fig. 7.5, for the same audio file used in Fig. 7.4. Note that the rate is expressed in bits and the distortion is a mean squared-error value. The number next to an operational point indicates the corresponding value of the codebook size M. The behaviour of the rate-distortion curve is as expected: as the codebook size is increased, the distortion diminishes while the rate grows.

The K-means clustering is initialized with reconstruction points selected at random, which can have an impact on the obtained clusters. For this reason, in the reported experiments the K-means clustering is repeated 10 times and the best solution is selected according to the overall minimum sum of distances of cluster members to centroids. To further mitigate the initialization effect, the process for computing every point in the curve is repeated 10 times, and the median values for rate and distortion are actually used.

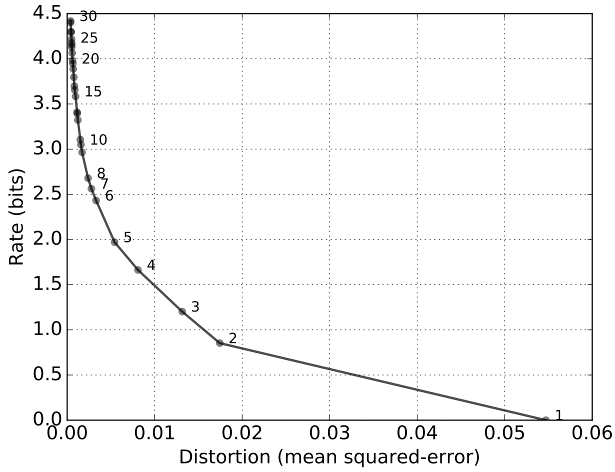


Figure 7.5: Operational rate-distortion curve obtained for the example of Fig. 7.4.

7.4 Experiments and results

Three different types of experiments are reported in this section aiming at assessing the usefulness of the proposed approach, using the dataset of *candombe* recordings introduced in Section 3.3. Firstly, the operational rate-distortion curves are used to qualitatively characterise drumming performances in terms of their overall complexity. Some possible implications of this method to the description of performance style and player expertise are also discussed. Then, the problem of estimating the number of different rhythmic patterns in a given performance is addressed within the rate-distortion framework. The solution investigated corresponds to selecting an operational point in the curve that adequately balances the rate-distortion trade-off. Finally, in the light of the previous experiments, the last problem addressed aims at identifying which one of the beats corresponds to the downbeat, without using any high-level information about the rhythm except for its four-beat structure. By comparing the rate-distortion curves of the different possible alignments of the four beats within the rhythm cycle, it turns out that the correct solution yields the less complex representation for a large part of the available recordings, thus allowing the automatic detection of the downbeat. The underlying rationale for the success of the method as well as its limitations are discussed and illustrated with examples.

7.4.1 Comparison of performance complexity

The operational rate-distortion trade-off will show a different behaviour depending on the complexity of the performance. Firstly, the rate-distortion curve is determined by the number of codevectors needed to properly encode the sequence. For instance, if there are several different rhythmic patterns played, then a small codebook size will not suffice to correctly describe the performance and will necessarily yield a high distortion value. Apart from that, there is also the issue of how well each group of patterns is represented by a single codevector. The amount of variability of the patterns within a certain group will also contribute to increase the distortion, even for the correct codebook size. For these reasons, when analysing different performances their rate-distortion curves will lie in different regions, simpler performances yielding lower rate-distortion values compared to the more complex ones. This is illustrated in the following experiment.

Experiment 7.1

Four different complete performances from the dataset were selected and classified by a music expert, namely Luis Jure, with regards to the overall complexity of the *piano* drum part. For each recording the low-frequency feature was extracted, the input sequence x^N was constructed using the beat/downbeat labels and the operational rate-distortion curve was computed varying the codebook size from 1 to 30. This is the standard procedure adopted for all the reported experiments. The resulting curves are depicted in Fig. 7.6, together with the corresponding input sequence x^N of each performance. From top to bottom, the input sequences are sorted in a decreasing order of complexity, according to the judgement of the music expert. Note that the same ordering is evidenced in the operational rate-distortion curves, the more complex performances indicated with darker lines.

There are many possible ways to characterise the rate-distortion curves and to summarize their behaviour into a single number. For instance, the distortion value for the codebook size $M = 1$ (that shall be denoted as D0), i.e. the point corresponding to zero rate, preserves the ordering of performance complexity. However, it ignores the behaviour of the curve for other codebook sizes. Another option is to compute the area under curve (AUC). To do that, the curve is extrapolated to estimate a cut-off point in the ordinates (with a polynomial fit considering the last 10 values) and the AUC is calculated using the numerical trapezoidal rule for approximating integrals. The AUC values obtained in this way are included in Fig. 7.6-left and are consistent with the qualitative ordering of the performances.

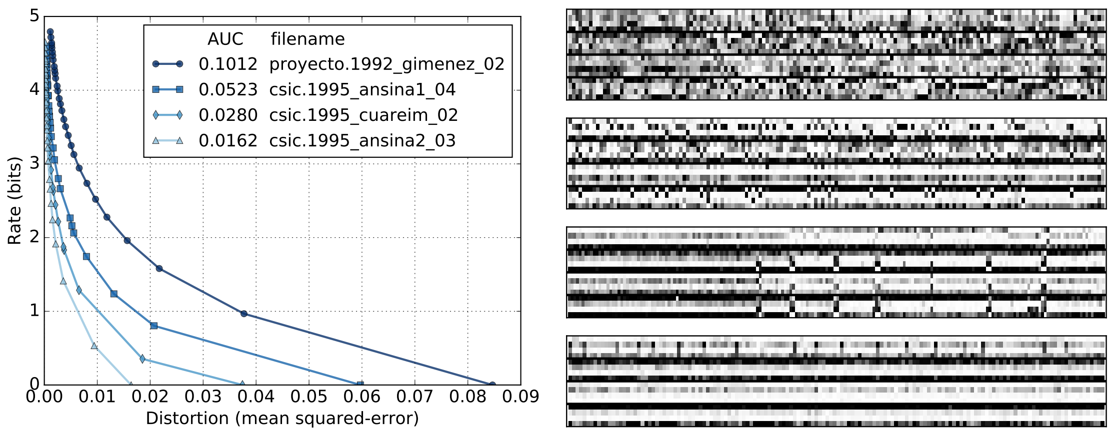


Figure 7.6: Rate-distortion curves (left) and input sequences (right) of Exp. 7.1.

The following description of the recordings provides some insight on the assessment proposed by the musicologist. The most complex recording is performed by the virtuoso *piano* drum player Eduardo “Malumba” Giménez, who belongs to the *Ansina* tradition. Although he has always been a highly accomplished player with a tendency of displaying feats of skill well above the average, in this particular take he deliberately intended to show a complex performance, as he explicitly stated once the recording was over. The next performance was recorded by Gustavo Oviedo, within the time period—of approximately twenty years—in which he was considered one of the best *piano* drum players of the *Ansina* style. The input sequence of the recording shows the irregular but well-balanced distribution of *base* and *repicado* patterns that was previously noted in Section 5.2. There are essentially four different rhythmic patterns, two *base* and two *repicado*, which also allow for some minor variations. The two remaining recordings are considerably less complex. The first one is in the *Cuareim* style, performed by noted *candombe* drummer Juan Silva. Arguably, the *piano* drum style in the *Cuareim* tradition is regarded as rather plain and with fewer *repicado* patterns compared to the *Ansina* style. After keeping the same *base* pattern for approximately the first one-third of the recording, it is the *piano* drum who calls for increasing the tempo by playing two *repicado* patterns. From this point onwards, the *base* pattern is simplified—probably due to the tempo increase—and the *repicado* pattern is played a few more times. The last recording features renowned *piano* drum player Eduardo “Cacho” Giménez, who has been an important and influential member of the *Ansina* style. However, in this recording session he limited himself to just playing only a single *base* pattern throughout the whole performance—which sometimes shows an ornamentation in the fourth beat—without any *repicado* patterns.

Experiment 7.2

It is reasonable to assume that the degree of complexity displayed in a performance is voluntarily controlled by the player, probably depending on the musical context. At the same time, the degree of complexity can also be associated with personal style and expertise. In order to illustrate these issues, different recordings of the same performers are considered in the following.

First, another performance by Eduardo “Malumba” Giménez is compared to the recordings of the previous experiment. Recall that this player produced the most complex performance of Fig. 7.6, which was in fact an exhibition of virtuosity. Yet, during the same session in 1992 he recorded a more conventional performance, whose rate-distortion curve is depicted in Fig. 7.7-left, along with the ones of the previous experiment. The evolution of the curve and its AUC value are clearly different from those of the first recording by the same player. Actually, it seems to be closer to the recording featuring Gustavo Oviedo (that ranked second in the previous experiment), which is quite in accordance with subjective assessment.

Then, a comparison is carried out considering all the performances of the *Ansina* style from the recording session of 1995. There are a total of 9 recordings, the *piano* drum played by Gustavo Oviedo in five of them and by Eduardo “Cacho” Giménez in the remaining four. The rate distortion curves and their

AUC values are presented in Fig. 7.7-right. Two groups of recordings are readily distinguishable, each one corresponding to a different performer. This indicates their personal styles were consistent and clearly different from each other during the whole recording session, which once again matches subjective assessment.

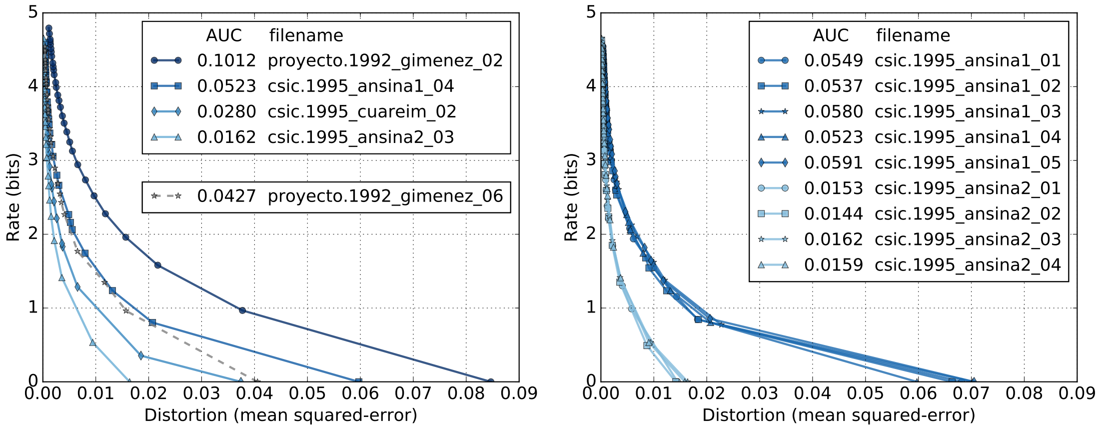


Figure 7.7: Rate-distortion curves of the first experiment with the addition of another recording by Eduardo "Malumba" Giménez (left) and comparison of all performances in the *Ansina* style from the recording session of 1995, corresponding to two different *piano* players (right).

7.4.2 Estimation of the number of rhythmic patterns

The next problem addressed is the automatic estimation of the number of rhythmic patterns in a given recording. This could be applied to the detection and classification of rhythmic patterns, performance style comparison and training of beat/downbeat tracking algorithms [219, 254]. Since the feature values are continuous and the rhythmic patterns may exhibit several variations within a recording, the problem can be regarded as finding a good compromise between a concise account of a given performance and a sufficiently precise description of its rhythmic patterns. Within the rate-distortion framework this corresponds to selecting a certain operating point of the trade-off. If a detailed representation is required, then the number of rhythmic patterns (i.e. the codebook size) has to be increased, at the expense of a necessarily longer performance description (i.e. higher rate). This can be posed as an optimization problem which can be solved using the discrete version of the Lagrangian method [225]. But it still requires one finds the optimal value for the Lagrange multiplier λ, a problem tackled in the following experiment.

Experiment 7.3

If a sufficiently large and representative training set is available, the optimal value for the Lagrange multiplier λ could be searched for, providing that it yields the correct number of rhythmic patterns for most of the data at hand. This kind of approach is illustrated in the following. A set of rhythmic patterns usually found in *candombe* performances was considered, and audio files that followed them were synthesized. To do that, the process described in Section 3.2 is applied, i.e. music scores with the rhythmic patterns were produced in a general purpose music engraving software language, adopting some conventions to represent the different types of strokes. Several sound samples of each type of stroke, recorded by a professional musician, were randomly selected by the synthesis program, which is able to interpret local accents and variations in dynamics.

The music scores of Fig. 7.8-left represent the six *piano* rhythmic patterns that were used in the experiment, comprising four *base* patterns and two *repicado* patterns. Lower and upper line represent hand and stick strokes respectively and the muffled strokes are indicated with a cross. Six audio files of the same length (180 rhythm cycles) were rendered by gradually incrementing the number of different patterns included, up to a uniform distribution of all of them. Fig. 7.8-right shows the of number of cycles per rhythmic pattern in each audio file.

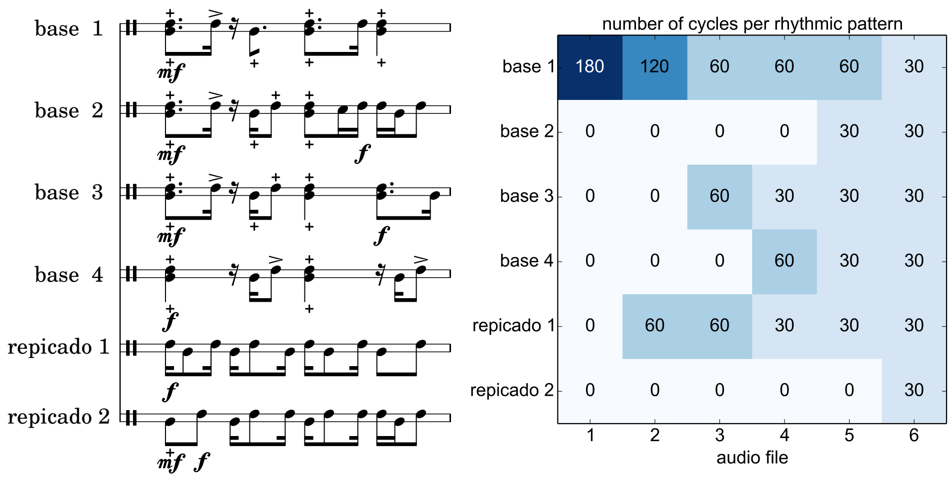


Figure 7.8: Rhythmic patterns used in the experiment (left) and number of cycles per rhythmic pattern in each audio file (right).

Then, the rate-distortion curves were computed and the discrete Lagrangian method was applied to them, i.e. the minimization in equation 7.9 was performed. For each curve, the λ values that yield the correct number of patterns were looked for, following a grid-search scheme. The grid of values considered is in the range $\lambda = [0.001, 0.05]$ with a step of 0.0001. Fig. 7.9 shows the rate-distortion curves, along with the extremes of the grid represented with a dashed line. The range of valid λ values for each audio file, i.e. the ones producing the correct number of patterns, is also indicated as a light greyed out region. If the extent of valid λ values among different files is considered, it turns out that the range $\lambda_{[1,6]} = [0.0058, 0.0099]$ yields the correct solution for all files, shown as a darker greyed out region. The next experiment tests this approach with real recordings.

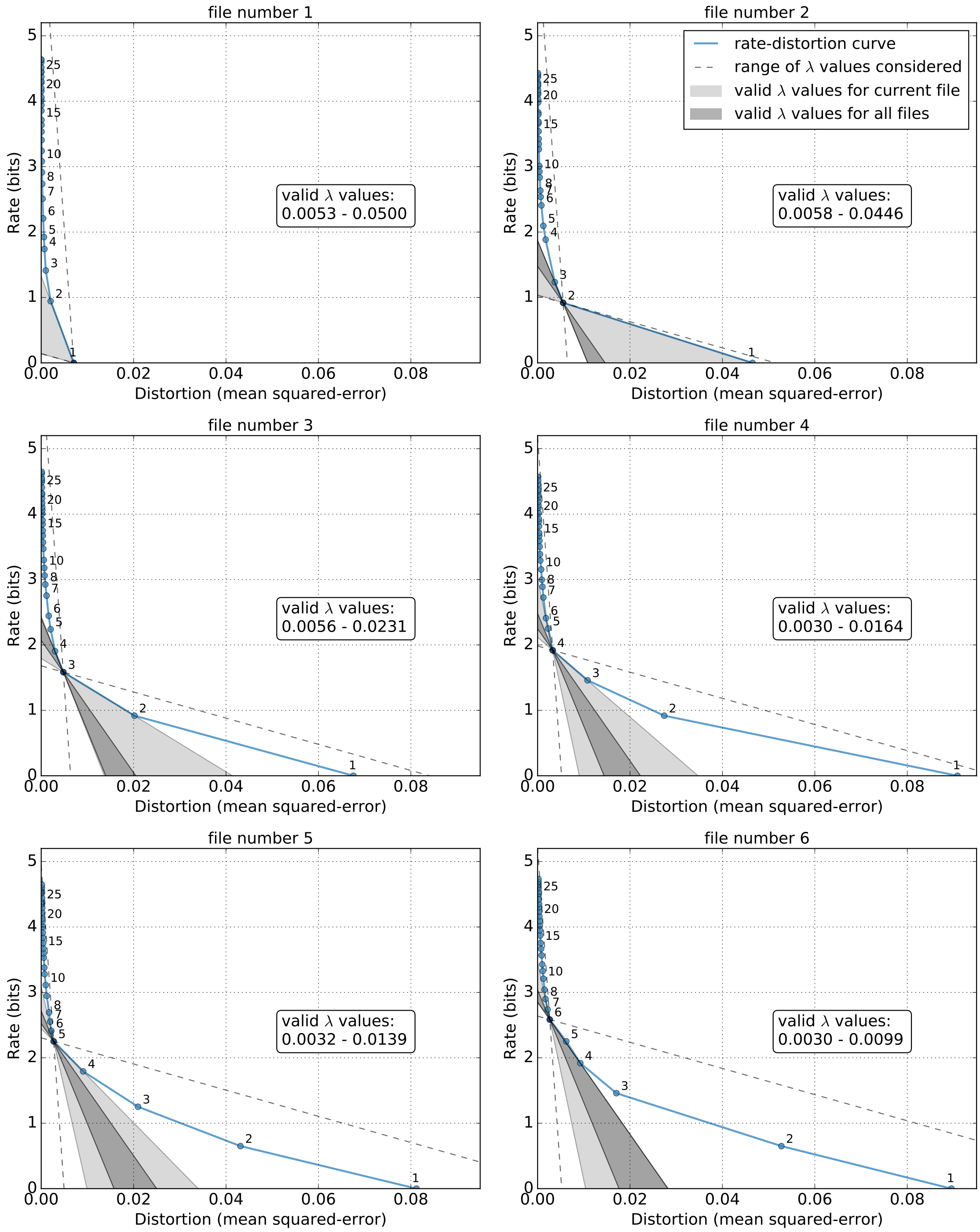


Figure 7.9: Rate-distortion curves for the synthetic audio files of Exp. 7.4. Extremes of the grid of λ values are depicted with dashed lines. The range of valid λ values for each file and the range in common for all files are shown as light and dark greyed out regions, respectively.

Experiment 7.4

This experiment tackles the estimation of the number of different rhythmic patterns in real recordings, considering the four complete performances introduced in Fig. 7.6. For this purpose, the discrete Lagrangian method is applied using a value of $\lambda^* = 0.00785$, which is the mean of the range $\lambda_{[1,6]}$. This is represented graphically in Fig. 7.10-left, as lines with slope $-1/\lambda^*$ intersecting each rate-distortion curve. The solutions obtained in this way suggest a number of patterns M of 6, 4, 3, and 2 for the recordings sorted in decreasing order of complexity. The encoding of each performance is presented in Fig. 7.10-right using the corresponding estimate of the number of patterns as the codebook size.

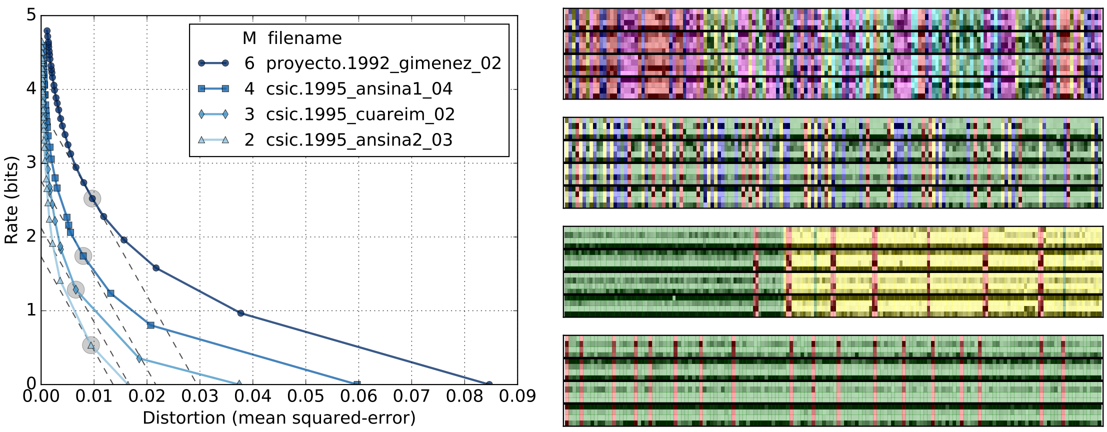


Figure 7.10: Estimation of the number of rhythmic patterns for the audio recordings of Figure 7.6. Dashed lines intersecting each rate-distortion curve (left) represent the discrete Lagrangian minimization applied. The resulting coding using the estimated number of patterns as the codebook size is depicted with colours in the feature maps (right).

Unlike the previous synthetic experiment there is no ground-truth in this case, but the validity of such estimations can be assessed in light of the description of the performances provided in Section 7.4.1. Recall that the less complex recording involved a single *base* pattern throughout the whole performance, which is sometimes modified by playing an ornamentation in the fourth beat. This description matches the resulting automatic coding, with one of the two rhythmic patterns being the plain *base* and the other corresponding to the ornamented one. For the next recording, the estimation of three different rhythmic patterns also agrees with the qualitative description, since the same *base* pattern is played during the first one third of the recording, and following a tempo increase—announced by *repicado* patterns—a variation of the *base* pattern is used until the end, with some *repicado* patterns in between. Then, also in accordance with the estimation, the following performance has essentially four different rhythmic patterns, two *base* and two *repicado*, that were described in detail in Section 5.2. Finally, the last recording is a puzzling one due to its virtuosity nature. There are several different rhythmic patterns that exhibit a lot of variation within each type. This makes it difficult to agree on the correct number of patterns, which are probably at least four. It can be noticed that similar consecutive patterns are grouped together and straight *repicado* patterns are separated from embellished *base* patterns.

It is interesting to examine the impact of the Lagrange multiplier into the estimation of the number of patterns of these recordings. To do that, instead of its mean value λ^*, the bounds of the range $\lambda_{[1,6]}$ are considered, that is $\lfloor\lambda\rfloor = 0.0058$ and $\lceil\lambda\rceil = 0.0099$. The estimations for the higher bound (i.e. the lowest slope) are 5, 4, 3, and 2, for the recordings in decreasing order of complexity, whereas if the lower bound is used (i.e. the highest slope) the estimated number of patterns are 7, 5, 3, and 2. The estimations for the last two recordings remain unchanged, while in the others the difference is only one pattern. This suggest certain robustness to the selection of the optimum λ value.

7.4.3 Downbeat detection

The last type of experiment recasts the downbeat detection task as a data compression problem. Assuming the location of beats is known, the aim is to identify which one of them corresponds to the downbeat. This is addressed by considering the different possible alignments of the four beats within the rhythm cycle. When rhythmic patterns of one-cycle length are considered, their alternation along the whole performance can give a hint on the location of the downbeat. In particular, the correct alignment will probably allow for a less complex description of the input sequence when compared to the misaligned options.

This is schematically illustrated in the example of Fig. 7.11. Consider the input vectors $x_j, j \in [1, N]$, one after the other as a single stream of features. To produce the input sequence x^N, they have to be assembled in groups of the length of a rhythm cycle, which is 16 *tatum* pulses in this case. Suppose there are only two different rhythmic patterns played, say a *base* and a *repicado* pattern (notated as b and r in Fig. 7.11). Therefore, the correct alignment—the one consistent with the downbeat—can be optimally represented with a codebook of only two codevectors. However, other alignments will produce rhythmic patterns that are combinations of the original ones, yielding *base-repicado* (br), *repicado-base* (rb) and *base-base* (bb) patterns. Therefore, a codebook of three codevectors is needed, leading to a more complex description of the input sequence.

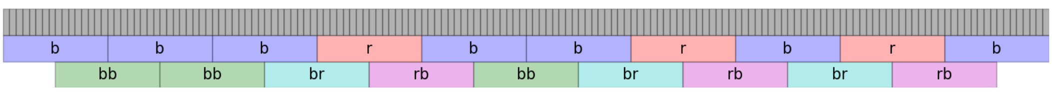


Figure 7.11: Diagram of two possible pattern alignments which imply different codebook sizes.

When the downbeat detection of an audio file is tackled, each beat of the four-beat rhythm cycle is alternatively considered as the downbeat, so four different alignments have to be evaluated. The different alignments are implemented as circular shifts of the feature map, starting from a shift of 0 beats (i.e. no shift) up to a shift of 3 beats. Larger shifts are redundant and therefore not considered—e.g.

a shift of 4 beats produces no shift. Then, operational rate-distortion curves are computed for each different alignment.

This is shown in Fig. 7.12 for four of the synthetic audio files of Section 7.4.2, involving 2, 3, 4 and 5 rhythmic patterns respectively. The complexity measures are also included in Fig. 7.12 for each one of the shifts, namely the area under curve (AUC), and the minimum value of the Lagrangian rate-distortion functional (Jmin). The discrete Lagrangian method is applied using a value of $\lambda^* = 0.00785$ (the mean of the range $\lambda_{[1,6]}$), and the codebook size M obtained in this way is also indicated. It can be seen that for all the audio files the correct alignment produces a curve that takes lower rate-distortion values compared to the shifted ones. Note that complexity measures also show this behaviour, and that even in those cases when the codebook size is the same the correct alignment yields a smaller distortion value, which indicates that each group of patterns is better represented by a single codevector.

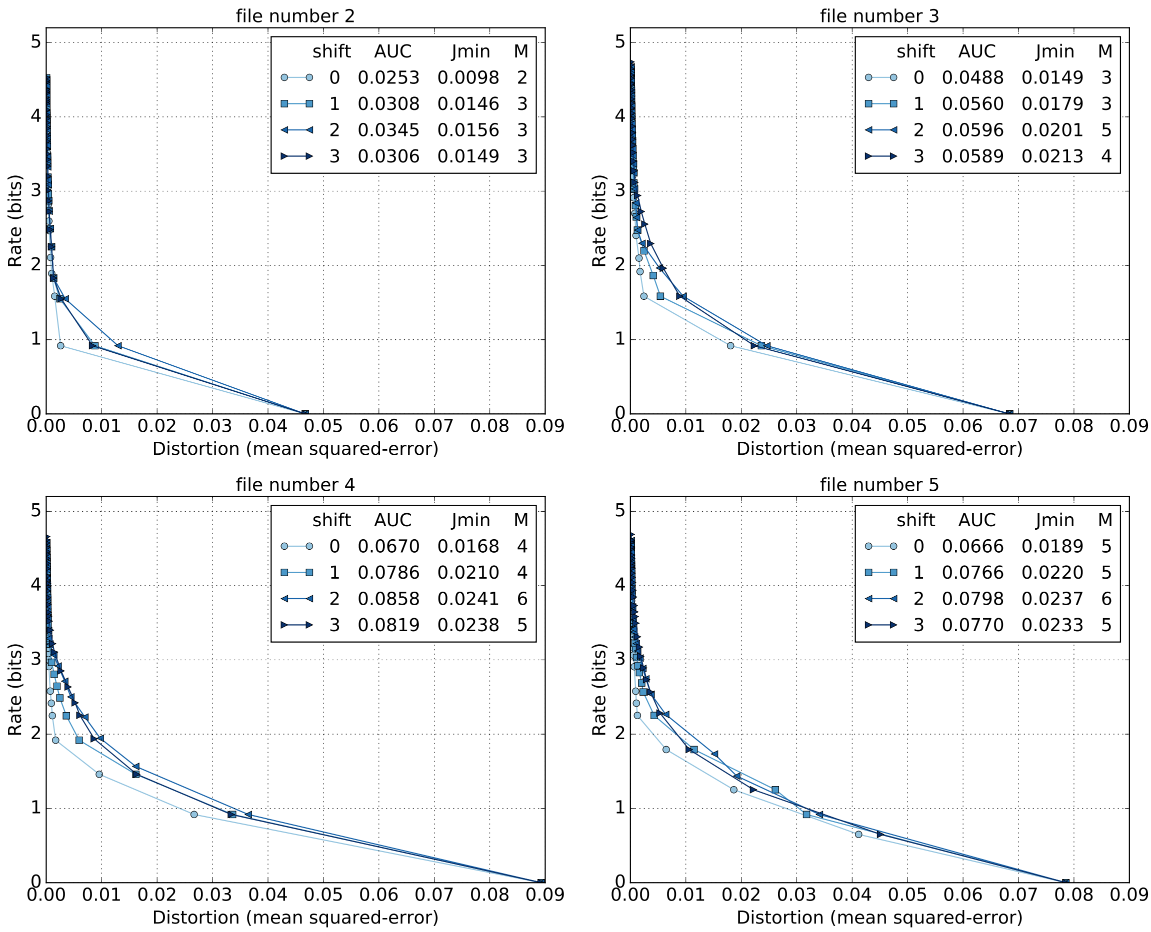


Figure 7.12: Downbeat detection analysis for three of the synthetic audio files introduced in Figure 7.8, involving 2, 3, 4 and 5 rhythmic patterns. The rate-distortion curves correspond to the four different possible alignments of the beats within the rhythm cycle.

The previous examples indicate that the downbeat could be identified by comparing the rate-distortion trade-off of the different alignments. Nevertheless, the rationale for this is the alternation of patterns in the recording and there may be some cases which fail to provide enough information for downbeat detection. To further illustrate this, the analysis of two real recordings is presented in Fig. 7.13. Recall the diagram of Fig. 7.11 in which the correct alignment yields the shortest codebook size. This situation corresponds to the first example of Fig. 7.13, which contains *base* and *repicado* patterns. The shifting of the feature map gives rise to a higher number of rhythmic patterns, so the complexity of the description needed to account for the performance is increased. Note that in this case both complexity measures favour the selection of the correct alignment.

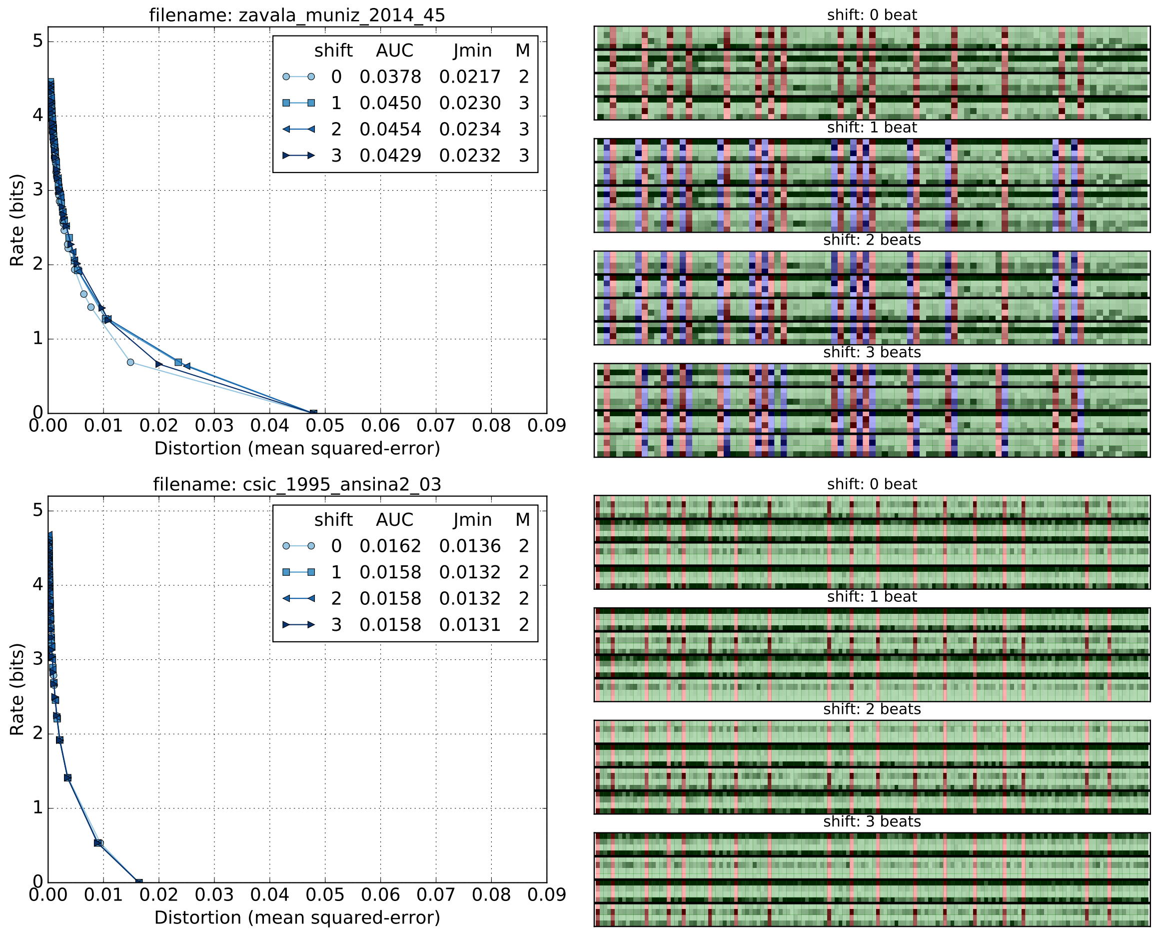


Figure 7.13: Downbeat detection analysis for two recordings of the dataset, one with *base* and *repicado* patterns (top) and another one with only a *base* pattern occasionally ornamented in the fourth beat (bottom). The rate-distortion curves (left) and the feature maps (right) correspond to the four different possible alignments of the beats within the rhythm cycle.

However, it is fairly obvious that if the performance contains a single pattern all alignments will be equivalent. Moreover, even in the case where there is more than one pattern the complexity of the different alignments may look all the same. For instance, in the second example of Fig. 7.13—the most simple of the recordings introduced in Fig. 7.6—the differences between the existing patterns are confined to a single beat. As previously noted, there is only one *base* pattern throughout the whole performance which sometimes shows an ornamentation in the fourth beat. Thus, shifting the patterns only relocates the ornamentation to a different beat. For this reason, the rate-distortion curves and measures for the different alignments provide no evidence to prefer one over the others.

This second example stresses the fact that the proposed downbeat detection approach requires not only the alternation of different rhythmic patterns but also that the differences between them span over the whole rhythmic cycle, as it happens in the first example of Fig. 7.13. In fact, if there were no differences between the rhythmic patterns for a certain beat during the entire recording, i.e. the four *tatums* of the beat always were articulated in the same way, then this beat would carry no information regarding the location of the downbeat and ambiguity would arise between different shifts. Nevertheless, note that differences between *base* and *repicado* patterns usually extend over the whole rhythmic cycle (see for instance the music scores of Fig. 7.8). Consequently, the method is likely to succeed for a performance that alternates the typical *base* and *repicado* patterns. On the other hand, if the performance is too simple, the downbeat may not be identified correctly. It is interesting to note that the degree of complexity of the performance could be estimated beforehand, even without knowing the location of the downbeat. The AUC or Jmin value for an arbitrary alignment could be used for that purpose, since their values are very similar for the different alignments (see Fig. 7.13). Actually, more reliable measures could be devised to assess the degree of confidence in the downbeat estimation taking into account the extent of the differences between the rhythmic patterns.

Experiment 7.5

In this experiment, the approach for downbeat detection based on information theory is tested on the dataset introduced in Section 3.3. For each recording the low-frequency feature was extracted and the beat/downbeat labels were used to render the four different alignments of the beats within the rhythm cycle. Then, for each different alignment the operational rate-distortion curve was computed and the complexity measures were calculated. The downbeat was estimated as the beat corresponding to the shift that minimizes the complexity measure.

The obtained results are presented in Table 7.1 for each file of the dataset, including the value of the two different complexity measures considered. The columns denoted as db1 and db2 are the number of the beat estimated as downbeat for the AUC and Jmin measures, respectively. A leave-one-out scheme was followed for determining the λ value for the discrete Lagrangian method, applying in each fold a process similar to that of Fig. 7.9. A correct detection means the downbeat is estimated as beat number 1, otherwise a failure is counted. The method attains an overall correct detection of 65.7% for the AUC, and 74.3% for the Jmin measure.

The rows of the table are sorted according to the AUC value in ascending order and those recordings in which the method fails for both measures are greyed out. Note that the first four recordings of Table 7.1—which belong to the same performer—are troublesome for the method (the AUC measure criterion fails in all of them, while the Jmin measure misses two). One of these recordings was already analysed in the second example of Fig. 7.13. Apart from having the lowest degree of complexity of the whole dataset (i.e. AUC value), all of them consist of a single pattern occasionally ornamented in the fourth beat, and as previously noted fail to provide enough information for downbeat detection. Both measures also fail in some other recordings, such as number 6, which exhibits only a single *base* pattern with a few simple variations. In this case, the patterns show virtually no differences at the first beat during the entire performance, thus leading to ambiguity in the selection of the downbeat. Something similar also happens with recording number 18, despite having a few *repicado* patterns.

It is interesting to note that for a large number of the recordings (22, 62.9%) the estimation of the downbeat is correct for both measures. As expected, several of these recordings belong the *Ansina* style and show an alternation of the typical *base* and *repicado* patterns that was shown to be advantageous for the method.

7.5 Discussion and conclusions

In summary, an approach based on the rate-distortion theory was proposed, which given an audio recording of a *candombe* performance computes a lossy representation that captures much of its underlying regularity but tolerates some amount of distortion. Thus, within a rate-distortion theory framework, the study of the trade-off between rate and distortion allows for the extraction of some relevant information about the performance.

Several experiments were conducted in order to assess the usefulness of the proposed approach when applied to a dataset of *candombe* drumming audio recordings. In particular, different performances were compared according to a measure of their overall complexity drawn from the operational rate-distortion curve, yielding results which roughly correspond to subjective judgement and correlate well with personal style and expertise. In addition, the estimation of the number of different rhythmic patterns in the recording was posed as the problem of selecting an operational point of the rate-distortion curve. The outcome of this method provided compact representations of the performances that are quite in accordance with manual analysis. Finally, the downbeat detection task was formulated as a data compression problem aiming at finding structure in the performance being analysed. To do that, the different possible alignments of the beats within the rhythm cycle were considered, and the one providing the most succinct representation—in terms of the rate-distortion trade-off—was selected as the downbeat. The method proved to be effective for a large part of the dataset, and the underlying rationale for its success as well as its limitations were discussed and illustrated with examples. Note that the Bayesian approach for rhythm analysis yielded downbeat detection results on the same dataset which are much better than those reported here, see Section 6.3. However, the algorithm tracks a rhythmic pattern, that is either based on a priori musical knowledge about the rhythm, or learned from the labelled database itself. The herein proposed method is less grounded on high-

num	db1	AUC	db2	Jmin	filename
1	2	0.0141	3	0.0141	csic.1995_ansina2_02
2	4	0.0146	1	0.0145	csic.1995_ansina2_01
3	2	0.0155	3	0.0158	csic.1995_ansina2_04
4	4	0.0158	1	0.0164	csic.1995_ansina2_03
5	2	0.0172	3	0.0166	proyecto.1992_lobo_01
6	3	0.0189	4	0.0201	proyecto.1992_pelado_05
7	1	0.0280	1	0.0241	csic.1995_cuareim_02
8	1	0.0312	1	0.0252	proyecto.1992_lobo_06
9	1	0.0338	1	0.0266	zavala.muniz.2014_41
10	4	0.0347	1	0.0289	csic.1995_cuareim_01
11	1	0.0360	1	0.0248	zavala.muniz.2014_51
12	1	0.0373	1	0.0295	csic.1995_cuareim_03
13	1	0.0376	1	0.0264	zavala.muniz.2014_52
14	1	0.0378	1	0.0285	proyecto.1992_pelado_01
15	1	0.0378	1	0.0257	zavala.muniz.2014_45
16	1	0.0391	1	0.0260	zavala.muniz.2014_46
17	1	0.0402	1	0.0269	zavala.muniz.2014_44
18	4	0.0415	4	0.0300	csic.1995_cuareim_05
19	4	0.0417	1	0.0317	csic.1995_cuareim_04
20	1	0.0427	1	0.0309	proyecto.1992_gimenez_06
21	3	0.0437	2	0.0303	proyecto.1992_magarinos_02
22	1	0.0495	1	0.0333	zavala.muniz.2014_42
23	1	0.0523	1	0.0327	csic.1995_ansina1_04
24	1	0.0537	1	0.0318	csic.1995_ansina1_02
25	1	0.0541	1	0.0337	zavala.muniz.2014_49
26	1	0.0549	1	0.0316	csic.1995_ansina1_01
27	1	0.0557	1	0.0348	zavala.muniz.2014_53
28	4	0.0571	4	0.0355	zavala.muniz.2014_48
29	1	0.0580	1	0.0339	csic.1995_ansina1_03
30	1	0.0591	1	0.0334	csic.1995_ansina1_05
31	1	0.0619	4	0.0364	zavala.muniz.2014_50
32	1	0.0642	1	0.0370	zavala.muniz.2014_54
33	1	0.0644	1	0.0369	zavala.muniz.2014_47
34	1	0.0767	1	0.0409	zavala.muniz.2014_55
35	4	0.1008	4	0.0462	proyecto.1992_gimenez_02
total	23	65.7%	26	74.3%	

Table 7.1: Downbeat detection results for the dataset from Section 3.3.

level information about the rhythm or in a training scheme, and constitutes a novel idea for tackling the downbeat detection problem, that could be combined with any other existing approach as another source of information.

A natural extension of the present work is taking into account the cost of describing the chosen model itself, as in an MDL approach. This implies the cost of describing the elements of the codebook of rhythmic patterns, which necessarily involves a quantization of their continuous feature values. Then, the cost of the model for coding the sequence of source symbols into the sequence of alphabet symbols has to be accounted for. For instance, if a Huffman coding for lossless data compression is adopted, then the cost of describing the coding trees has to be taken into account. This is illustrated in the rate-distortion curves of Fig. 7.14 for a real performance. The cost of describing the rhythmic patterns using a uniform quantizer of 4 levels and the cost of describing the Huffman coding trees is depicted, along with the rate of the encoded sequence. The total sum of the rates, also shown, could be used for the type of analysis previously described. As can be seen, the cost of coding the Huffman trees is almost negligible, but the cost of describing codebook elements influences the total rate, in particular as the codebook size increases. In spite of that, some simulations conducted yielded very similar results compared to the ones reported, but this should be further studied in future work. Besides, other information theory frameworks for model selection will be tested in future work. The application of the proposed approach for downbeat detection and structure analysis to other types of music which also exhibit repeated rhythmic patterns (e.g. Afro-Brazilian and Afro-Cuban) is one of the most promising future research strands.

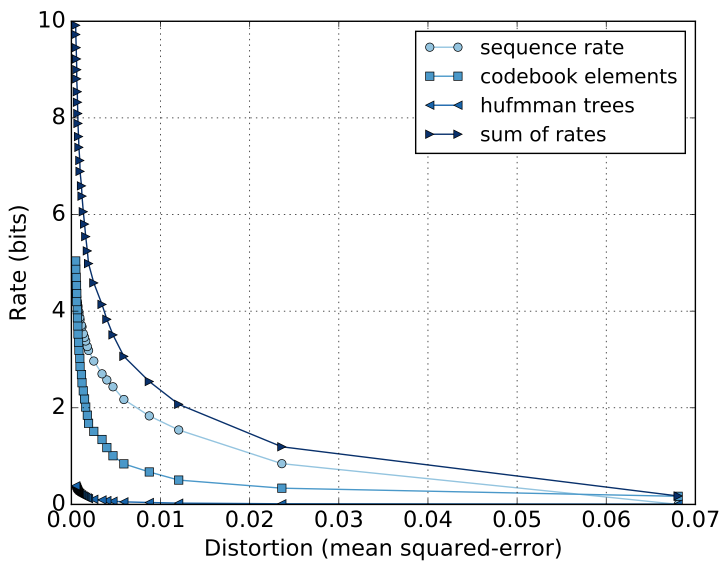


Figure 7.14: Operational rate-distortion curves including the cost of the model.

Chapter 8

Conclusions

8.1 Summary and conclusions

The following is a summary of the thesis, which also highlights the key results, main contributions and conclusions of the work. On the whole, the dissertation aimed to build computational methods for the analysis of rhythm from audio recordings of percussion music. A domain–specific and culture–aware approach was favoured, oriented towards the Afro-Uruguayan *candombe* drumming.

Definition of a domain–specific and culture–aware perspective

The dissertation offered a comprehensive overview of the historical, social and cultural context in which *candombe* drumming is embedded. In addition, a description of the rhythm and its performance practices was also provided. This showed some of the traits that link *candombe* to other Afro-Atlantic music traditions and that consequently differentiate it from most usual rhythmic structures found in the so–called 'Western' music. In brief, the *clave* pattern divides the rhythmic cycle irregularly, the rhythmic pattern that defines the pulse—that of the *chico* drum—usually does not articulate the beat and has instead an accent on the second *tatum*, and some other strong phenomenological accents—from the *piano* and *repique* drums—are displaced with respect to the metric structure.

Creation of music collections and annotated datasets

One of the specific contributions of the thesis was to build annotated music datasets of *candombe* drumming suitable for research on computational rhythm analysis. The musical setting considered was small-size *candombe* ensembles, of three to five drums. An important amount of work was dedicated to collecting and labelling audio recordings. Besides, recording sessions were conducted with the aim to produce an audio–visual dataset of *candombe* performances, that could serve both documentary and research purposes. A detailed description of the tasks, requirements and technical challenges involved in producing the audio-visual records was provided. The following contributions are highlighted as a result of this work.

- A dataset of annotated recordings for beat and downbeat tracking was publicly released, being the first resource of this type for *candombe* [219].

- An audio-visual dataset of *candombe* performances with multi–track audio recordings and a multiple-camera set-up was obtained [197,255]. The dataset includes different types of annotations, such as metrical information (beat and downbeat), location of onsets and sections. Apart from its documentary value, it is an important resource for research, that is being used outside the scope of this dissertation for conducting studies on interpersonal entrainment in music performance [251][1] and will be also made publicly available.

Discovery and analysis of rhythmic patterns

Part of the dissertation focused on the discovery and analysis of rhythmic patterns from audio recordings. Efforts have been put in the formulation of relevant analysis problems in the context of *candombe* drumming.

One of the type of problems addressed involves the study of rhythmic patterns that span over the whole rhythmic cycle. For this purpose, a representation in the form of a map of rhythmic patterns was devised, which is based on spectral onset features in different frequency bands. When applied to the analysis of a recording, it allows the inspection of the similarities and differences of the rhythmic patterns used throughout the performance. Some experiments were carried out to study the extended hypothesis that the *piano* drum patterns show stylistic differences related to both personal and traditional performing styles. The unsupervised grouping of rhythmic patterns from different recordings showed certain correlation with the traditional styles, but a predominantly tendency to be clustered by performer. It was even possible to recognized who was playing the *piano* drum in a certain recording from a reduced group of performers. Certainly, additional information should be considered to properly characterize *piano* drum performing styles, but the experiments are just indicative of the potential usefulness of these kind of tools.

In addition, a pattern discovery problem was formulated to study the characteristics of the *clave* pattern. It is based on the fact that, instead of a single timeline pattern as in other Afro–Latin–American musics, the *clave* pattern in *candombe* allows for several different types and variants. To that end, a method was proposed in which the rhythm cycles in a given recording where the *clave* pattern is played are identified using automatic onset detection and sound classification. After that, the identified rhythmic patterns are clustered according to their similarity, thus illustrating all the different ways in which the *clave* was played. Indeed, some experiments showed that almost all the relevant *clave* variants reported in the musicological literature were collected with this method from a relatively small dataset of recordings. The comparison of all the collected *clave* patterns reveals some invariances that stand out as an underlying structure of the rhythm.

[1]Within the Interpersonal Entrainment in Music Performance (IEMP) project. `https://musicscience.net/projects/iemp/`

Micro-rhythmic properties of the rhythm

Turning now to the other type of analysis tackled, the dissertation also studied the micro–rhythmical properties of the drumming patterns in *candombe*. According to the obtained results, the *chico* pattern exhibits a contraction of the inter–onset intervals, that actually does not fit current musicological descriptions. Additionally, the deviation of the *repique* primary pattern with respect to the four pulses of the beat towards a triplet feeling, was somehow confirmed but more precisely characterized. Overall, the analysis of several recordings revealed the systematic use of micro–rhythmical deviations in the patterns of *candombe*, indicating that micro–timing is a structural component of its rhythm. This can be considered as an evidence of the existence of a sort of "swing" characteristic of *candombe* drumming. These findings, while preliminary, suggest that *candombe* rhythm has an isochronous grid of beats, that exhibits an uneven subdivision structure, following a short-short-short-long (SSSL) pattern. To the best of our knowledge this is the first systematic study and characterization of the micro–rhythmic structure of *candombe* drumming from a dataset of multi-track audio recordings.

Automatic meter inference and tracking from audio recordings

The rest of the dissertation was devoted to the automatic inference and tracking of the metric structure from audio recordings. The complexities and characteristics of *candombe* drumming were appropriate to push the boundaries of the state of the art in automatic rhythm analysis in MIR. Indeed, the rhythm proved to be very challenging for most of the state–of–the–art methods for beat and downbeat tracking. For this reason, a supervised scheme for rhythmic pattern tracking was proposed in this thesis [219]—based on the Bayesian approach for rhythmic analysis [293]—and a software implementation was publicly released. It aims at finding the metric structure from an audio signal, including the phase of beats and downbeats, by tracking a rhythmic pattern in the low–frequency band. Experiments employing both rhythmic patterns based on musical knowledge and others learned from the annotated data showed that the proposed algorithm can estimate the beat and downbeat positions correctly for most of the files in the dataset. However, in its present form, the main drawback of the proposed algorithm is its limited ability to properly deal with tempo changes. A more elaborate algorithm [174], recently published, also based on the Bayesian approach but able to track tempo variations, showed a virtually perfect performance when trained and tested with the dataset of *candombe* recordings. Therefore, the present work gives additional evidence of the generalizability of the Bayesian approach to complex rhythms from different music traditions. Moreover, the obtained results are very encouraging and allow us to confidently tackle other problems and applications that rely on automatic tracking of the metric structure from *candombe* recordings.

Finally, the downbeat detection task was formulated as a data compression problem aiming at finding structure in the performance being analysed. To do that, the different possible alignments of the beats within the rhythm cycle were considered, and the one providing the most succinct representation—in terms of the rate–distortion trade–off—was selected as the downbeat. The method proved to be effective for a large part of the dataset, and the underlying rationale for its success as well as its limitations were discussed and illustrated with examples. This is a novel idea for tackling the downbeat detection problem, that could be combined with other existing methods and applied to other types of music that also exhibit repeated rhythmic patterns. Additionally, it turned out that the obtained description was well suited for addressing other related tasks, namely assessment of performances' complexity and estimation of the number of different rhythmic patterns in a given recording.

8.2 Future work perspectives

There are several directions for future research based on the work presented in this dissertation. The thesis attempted to identify research problems in rhythm analysis of *candombe* drumming, which when addressed with the technologies developed could lead to practical results that were musically relevant and useful. In some of these problems, however, this dissertation has barely scratched the surface.

For example, the front–end used for the characterization of the rhythmic patterns should be further investigated, in order to replace the simplistic spectral flux based audio feature with a more appropriate alternative. Probably the most promising strand to follow is to learn the relevant features directly from spectral representations of the audio signals, by using, for instance, Convolutional Neural Networks [96, 143]. The classification of the type of stroke for each articulated pulse is also envisioned as a possible improvement of the proposed techniques for some particular problems, such as pattern discovery and retrieval.

In addition, all the experiments conducted on rhythmic pattern analysis and discovery, as well as the study of the micro–rhythmic structure, should be explored further on larger datasets of audio recordings. In this way, some of the findings that were suggested in this work could be rigorously tested to produce significant musicological conclusions. Fortunately, increasing the datasets becomes more feasible now, given the performance attained by the beat and downbeat tracking algorithms adapted to *candombe* drumming, because manual annotations of the metric structure seem to be not longer needed.

In fact, the availability of reliable algorithms for metric analysis opens up the possibility for addressing several other music research questions. In future work, instead of small–size *candombe* ensembles, large percussion groups will be tackled, such as the *comparsas* that parade in Carnival. Analysis of their different performance styles, considering for instance tempo and loudness evolution over time, are interesting topics for future investigation. Similarly, the beat and downbeat tracking tools should be extended to deal with groups that include, apart from *candombe* drums, other musical instruments.

One of the most appealing treats of *candombe* is the role of the *repique* drum, which has the greatest degree of freedom and is the main responsible of musical variety. Therefore, one of our main goals for future research is to apply computational tools to study improvisation in *candombe* following recent studies in the musicological field carried out by close collaborators [158].

There are several other open questions related to the work presented in this dissertation that still have to be answered. The analysis of the micro-rhythmic deviations of the different ensemble parts of a *candombe* performance leaves a lot of room for investigation. For instance, it is very relevant to try to figure out if such tiny deviations are actually perceptible, and if there is any cultural bias in the ability of perceiving them or not. It is also interesting to study the very tight synchronization that the *candombe* drummers seem to achieve in their performances, as well as to characterize their timing discrepancies and other music entrainment processes involved. Some of this issues are actually being addressed in an ongoing research project about music entrainment [251].

Integration of developed algorithms into practical applications for music learning and performing is one of the most appealing directions for future work.

This dissertation strived to open up new paths for research in the application of computational tools for the analysis of *candombe* drumming. We hope that the future directions discussed here could be motivating for other researchers to join.

Appendix A

List of performers

The performers acknowledged hereafter took part in the recordings of the dataset for beat and downbeat tracking [219].

Mariano Barroso
Eduardo ‘Cacho’ Giménez
Eduardo ‘Malumba’ Giménez
Francisco Giménez
José Luis Giménez
Jorge ‘Foqué’ Gómez
José Pedro ‘Perico’ Gularte
Luis ‘Pocholo’ Maciel
Julio Magariños
Raúl ‘Neno’ Magariños
[. . .] Magariños
Javier ‘Cerdo’ Martirena
Wilson Martirena
Eduardo ‘Tierra’ Nilo
Sergio Ortuño
Fernando ‘Lobo’ Núñez
Edinson ‘Palo’ Oviedo
Gustavo Oviedo
[. . .] Pintos
Luis ‘Mocambo’ Quiroz
Rodolfo ‘Pelado’ Rodríguez
Fernando ‘Hurón’ Silva
Juan Silva
Raúl Silva
Waldemar ‘Cachila’ Silva
Héctor Manuel Suárez

Appendix B

Software tools

Software tools and experiments were implemented in Python, using Numpy, Scipy, Matplotlib and Scikit-learn libraries. Music examples were typeset using LilyPond, which was also applied together with Csound for the synthesis of test audio signals. In order to leverage the software provided by other researchers also Matlab and R were used. All the operating systems used were based on GNU/Linux.

Appendix C

Synthetic performances scores

example 1


10
12
14
16
18


20
22
24
26
28


30
32
34

example 2


9
11
13
15
17


19
21
23
25
27


29
31
33
35
37

Bibliography

[1] Samer Abdallah, Nicolas Gold, and Alan Marsden. Analysing symbolic music with probabilistic grammars. In David Meredith, editor, *Computational Music Analysis*, pages 157–189. Springer International Publishing, 2016.

[2] M. Aboy, R. Hornero, D. Abasolo, and D. Alvarez. Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE Transactions on Biomedical Engineering*, 53(11):2282–2288, Nov 2006.

[3] Laura Adinolfi and Carina Erchini. El conventillo Mediomundo: materialidad e inmaterialidad en el barrio Sur. In *Almanaque del Banco de Seguros del Estado*, pages 131–138. BSE, Montevideo, Uruguay, 2007.

[4] V. Kofi Agawu. 'Gi Dunu,' 'Nyekpadudo,' and the Study of West African Rhythm. *Ethnomusicology*, 30(1):64–83, 1986.

[5] Coriún Aharonián. La música del tamboril. Primera parte. In *Músicas populares del Uruguay*, pages 83–91. Comisión Sectorial de Educación Permanente, Universidad de la República (2007), Montevideo, Uruguay, 1991.

[6] Coriún Aharonián. La música del tamboril. Segunda parte. In *Músicas populares del Uruguay*, pages 92–98. Comisión Sectorial de Educación Permanente, Universidad de la República (2007), Montevideo, Uruguay, 2000.

[7] Coriún Aharonián. El candombe, los candombes. In *Músicas populares del Uruguay*, pages 99–112. Comisión Sectorial de Educación Permanente, Universidad de la República (2007), Montevideo, Uruguay, 2007.

[8] Coriún Aharonián. La murga, lo murgustico. In *Músicas populares del Uruguay*, pages 113–123. Comisión Sectorial de Educación Permanente, Universidad de la República (2007), Montevideo, Uruguay, 2007.

[9] Milita Alfaro. *Carnaval. Una historia social de Montevideo desde la perspectiva de la fiesta. Primera Parte. El Carnaval heroico (1800–1872).* Editorial Trilce, Montevideo, Uruguay, 1991.

[10] Milita Alfaro. *Carnaval. Una historia social de Montevideo desde la perspectiva de la fiesta. Segunda Parte. Carnaval y Modernización. Impulso y freno del disciplinamiento (1873–1904).* Editorial Trilce, Montevideo, Uruguay, 1998.

[11] Milita Alfaro. *Memorias de la Bacanal. Vida y milagros del Carnaval montevideano (1850-1950).* Ediciones de la Banda Oriental, Montevideo, Uruguay, 2008.

[12] Milita Alfaro. *Jaime Roos: El Montevideano. Vida y obra.* Editorial Planeta, Montevideo, Uruguay, 2017.

[13] Milita Alfaro and José Cozzo. *Mediomundo, Sur, conventillo y después.* Editorial Medio&Medio, Montevideo, Uruguay, 2008.

[14] Cyril Allauzen, Maxime Crochemore, and Mathieu Raffinot. Factor oracle: A new structure for pattern matching. In *Proc. of the 26th Conference on Current Trends in Theory and Practice of Informatics*, pages 295–310, Milov, Czech Republic, Nov. 1999.

[15] Olavo Alén. Rhythm as duration of sounds in Tumba Francesa. *Ethnomusicology*, 39(1):55–71, 1995.

[16] George R. Andrews. *Afro-Latin America, 1800-2000.* Oxford University Press, New York, USA, 2004.

[17] George R. Andrews. Remembering Africa, Inventing Uruguay: Sociedades de Negros in the Montevideo Carnival, 1865-1930. *Hispanic American Historical Review*, 4(87):693–726, 2007.

[18] George R. Andrews. *Blackness in the White Nation: A History of Afro-Uruguay.* The University of North Carolina Press, Chapel Hill, USA, 2010.

[19] Dana Angluin and Carl H. Smith. Inductive inference: Theory and methods. *ACM Computing Surveys*, 15(3):237–269, September 1983.

[20] Willie Anku. Principles of rhythm integration in African drumming. *Black Music Research Journal*, 17(2):211–238, 1997.

[21] Willie Anku. Circles and time: A theory of structural organization of rhythm in African music. *Music Theory Online*, 6(1), 2010.

[22] Darío Arce Asenjo. El desfile de las Llamadas como ritual conmemorativo. In Gustavo Goldman, editor, *Cultura y sociedad afro-rioplatense*, pages 127–150. Ediciones Perro Andaluz, Montevideo, Uruguay, 2008.

[23] Felipe Arocena. La contribución de los inmigrantes en Uruguay. *Papeles del CEIC. International Journal on Collective Identity Research*, (47), September 2009.

[24] Felipe Arocena and Sebastián Aguiar. *Multiculturalismo en Uruguay.* Editorial Trilce, Montevideo, Uruguay, 2007.

[25] G. Assayag and S. Dubnov. Using factor oracles for machine improvisation. *Soft Computing*, 8(9):604–610, September 2004.

[26] G. Assayag, S. Dubnov, and O. Delerue. Guessing the composer's mind: Applying universal prediction to musical style. In *Proc. of the International Computer Music Conference*, pages 496–499, Beijing, China, Oct. 1999.

[27] Lauro Ayestarán. La Música Negra. La Música en el Uruguay, 1953.

[28] Lauro Ayestarán. *El folklore musical Uruguayo*. Editorial Arca, Montevideo, Uruguay, 1967.

[29] Lauro Ayestarán. La 'conversación' de tamboriles. *Revista Musical Chilena, Santiago de Chile*, (101), 1967.

[30] Lauro Ayestarán. El tamboril afrouruguayo. *Boletín Interamericano de Música, Washington*, (68), 1968.

[31] Lauro Ayestarán, Flor de María Rodríguez, and Alejandro Ayestarán. *El tamboril y la comparsa.* Editorial Arca, Montevideo, Uruguay, 1990.

[32] A. Barron, J. Rissanen, and Bin Yu. The minimum description length principle in coding and modeling. *IEEE Transactions on Information Theory*, 44(6):2743–2760, October 1998.

[33] José Pedro Barrán. *Historia de la sensibilidad en el Uruguay. La cultura "bárbara" (1800-1860). El disciplinamiento (1860-1920).* Ediciones de la Banda Oriental, Montevideo, Uruguay, 2008.

[34] Bruce Bartlett and Jenny Bartlett. *Recording Music on Location.* Focal Press, Oxford, 2nd edition, 2007.

[35] J.P. Bello. Measuring structural similarity in music. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7):2013–2025, Sept 2011.

[36] Juan P. Bello, Robert Rowe, Carlos Guedes, and Godfried Toussaint. Five perspectives on musical rhythm. *Journal of New Music Research*, 44(1):1–2, 2015.

[37] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, 13(5):1035–1047, Sept 2005.

[38] Fernando Benadon. Slicing the beat: Jazz eighth-notes as expressive microrhythm. *Ethnomusicology*, 50(1):73–98, 2006.

[39] Fernando Benadon. Near-unisons in Afro-Cuban ensemble drumming. *Empirical Musicology Review*, 11(2), 2016.

[40] Toby Berger. Rate distortion theory and data compression. In International Centre for Mechanical Sciences, editor, *Advances in Source Coding*, volume 166, chapter 1, pages 3–39. Springer-Verlag Wien, New York, USA, 1975.

[41] Toby Berger and Jerry D. Gibson. Lossy source coding. *IEEE Trans. Information Theory*, 44(6):2693–2723, 1998.

[42] D. E. Berlyne. The new experimental aesthetics. In D. E. Berlyne, editor, *Studies in the new experimental aesthetics: steps toward an objective psychology of aesthetic appreciation,*, pages 1–25. Hemisphere Pub. Corp., 1974.

[43] Thierry Bertin-Mahieux, Daniel P. W. Ellis, Brian Whitman, and Paul Lamere. The million song dataset. In *Proc. of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, pages 591–596, Miami, Florida, USA, Oct. 2011.

[44] Jeff Bilmes. A model for musical rhythm. In *Proceedings of the 1992 International Computer Music Conference*, pages 207–210, San Francisco, International Computer Music Association, 1992.

[45] Jeff Bilmes. Timing is of the essence: Perceptual and computational techniques for representing, learning, and reproducing expressive timing in percussive rhythm. Master's thesis, MIT, Cambridge, MA, 1993.

[46] Frédéric Bimbot. Towards an Information-Theoretic Framework for Music Structure. In Meinard Müller, Elaine Chew, and Juan Pablo Bello, editors, *Computational Music Structure Analysis (Dagstuhl Seminar 16092)*, volume 6, pages 167–168. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2016.

[47] Rachel M. Bittner, Justin Salamon, Mike Tierney, Matthias Mauch, Chris Cannam, and Juan Pablo Bello. MedleyDB: A multitrack dataset for annotation-intensive MIR research. In *Proc. of the 15th International Society for Music Information (ISMIR 2014)*, pages 155–160, Taipei, Taiwan, Oct. 2014.

[48] S. Böck, F. Krebs, and G. Widmer. A multi-model approach to beat tracking considering heterogeneous music styles. In *Proc. of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, pages 603–608, Taipei, Taiwan, Oct. 2014.

[49] Sebastian Böck, Filip Korzeniowski, Jan Schlüter, Florian Krebs, and Gerhard Widmer. madmom: a new Python Audio and Music Signal Processing Library. In *Proceedings of the 24th ACM International Conference on Multimedia*, pages 1174–1178, Amsterdam, The Netherlands, 10 2016.

[50] Sebastian Böck, Florian Krebs, and Markus Schedl. Evaluating the online capabilities of onset detection methods. In *Proc. of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, pages 49–54, Porto, Portugal, Aug. 9-13 2012.

[51] Sebastian Böck, Florian Krebs, and Gerhard Widmer. Joint beat and downbeat tracking with recurrent neural networks. In *Proc. of the 17th International Society for Music Information Retrieval Conference (ISMIR 2016)*, pages 255–261, New York, USA, Aug. 2016.

[52] Sebastian Böck and Markus Schedl. Enhanced beat tracking with context-aware neural networks. In *Proceedings of the 14th International Conference on Digital Audio Effects (DAFx-11)*, Paris, France, Sep. 19-23 2011.

[53] Sebastian Böck and Gerhard Widmer. Maximum filter vibrato suppression for onset detection. In *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13)*, Maynooth, Ireland, Sep. 2-5 2013.

[54] Peter Boot, Anja Volk, and W. Bas de Haas. Evaluating the role of repeated patterns in folk song classification and compression. *Journal of New Music Research*, 45(3):223–238, 2016.

[55] Alex Borucki. Tensiones raciales en el juego de la representación. Actores afro en Montevideo tras la fundación republicana (1830-1840). In Gustavo Goldman, editor, *Cultura y sociedad afro-rioplatense*, pages 243–267. Ediciones Perro Andaluz, Montevideo, Uruguay, 2008.

[56] Alex Borucki. *Abolicionismo y tráfico de esclavos en Montevideo tras la fundación republicana (1829–1853)*. Biblioteca Nacional (MEC) y Facultad de Humanidades y Ciencias de la Educación (Universidad de la República), Montevideo, Uruguay, 2009.

[57] Alex Borucki. The Slave Trade to the Río de la Plata (1777–1812): Trans-Imperial Networks and Atlantic Warfare. *Colonial Latin American Review*, 20(1):81–107, Apr. 2011.

[58] Alex Borucki. *From Shipmates to Soldiers: Emerging Black Identities in the Río de la Plata*. University of New Mexico Press, Albuquerque, 2015.

[59] Alex Borucki, Karla Chagas, and Natalia Stalla. *Esclavitud y trabajo. Un estudio sobre los afrodescendientes en la frontera uruguaya (1835–1855)*. Pulmón ediciones, Montevideo, Uruguay, 2004.

[60] Roberto Bracco, José María López Mazz, Beatriz Orrego Rojas, Nicolás Batalla, Rodrigo Bongiovanni, and Magdalena Gutiérrez. *Esclavitud y afrodescendientes en Uruguay: Una mirada desde la antropología*. Universidad de la República, Facultad de Humanidades y Ciencias de la Educación, Montevideo, Uruguay, 2012.

[61] Marisa Bucheli and Wanda Cabella. *El perfil demográfico y socioeconómico de la población uruguaya según su ascendencia racial*. Instituto Nacional de Estadística, INE, Montevideo, Uruguay, 2006.

[62] James Burns. Rhythmic archetypes in instrumental music from Africa and the diaspora. *Music Theory Online*, 16(4), 2010.

[63] Matthew W. Butterfield. Why do jazz musicians swing their eighth notes? *Music Theory Spectrum*, 33(1):3–26, 2011.

[64] Wanda Cabella and Mathías Natan. *Iguales y diferentes. La composición de la población uruguaya desde la perspectiva étnico-racial*, volume 20 of *Libro de los Bicentenarios*. Comisión del Bicentenario, IMPO, Montevideo, Uruguay, 2013.

[65] Wanda Cabella, Mathías Natan, and Mariana Tenenbaum. La población afro-uruguaya en el Censo 2011. In Juan José Calvo, editor, *Atlas sociodemográfico y de la desigualdad del Uruguay, Fascículo 2*. Facultad de Ciencias Sociales, Universidad de la República / Ediciones Trilce, Montevideo, Uruguay, 2013.

[66] Chris Cannam, Christian Landone, Mark B. Sandler, and Juan Pablo Bello. The sonic visualiser: A visualisation platform for semantic descriptors from musical signals. In *Proc. of the 7th International Conference on Music Information Retrieval (ISMIR 2006)*, pages 324–327, Victoria, Canada, Oct. 2006.

[67] Sofia Cavaco and Hugo Almeida. Automatic cymbal classification using non-negative matrix factorization. In *19th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 468–471, April 2012.

[68] Ò. Celma, Perfecto Herrera, and Xavier Serra. Bridging the music semantic gap. In *Workshop on Mastering the Gap: From Information Extraction to Semantic Representation*, volume 187, Budva, Montenegro, 11-06-2006 2006. CEUR, CEUR.

[69] Karla Chagas and Natalia Stalla. *Recuperando la memoria. Afrodescendientes en la frontera uruguayo brasileña a mediados del siglo XX*. Ministerio de Educación y Cultura, Montevideo, Uruguay, 2009.

[70] Gregory J. Chaitin. On the length of programs for computing finite binary sequences. *Journal of the Association for Computer Machinery (ACM)*, 13(4):547–569, October 1966.

[71] J.C. Chasteen. *National Rhythms, African Roots: The Deep History of Latin American Popular Dance*. University of New Mexico Press, 2004.

[72] Song Xi Chen. Beta kernel estimators for density functions. *Computational Statistics & Data Analysis*, 31(2):131–145, 1999.

[73] R. Cilibrasi and P. M. B. Vitányi. Clustering by compression. *IEEE Transactions on Information Theory*, 51(4):1523–1545, April 2005.

[74] Rudi Cilibrasi, Paul Vitányi, and Ronald De Wolf. Algorithmic clustering of music based on string compression. *Computer Music Journal*, 28(4):49–67, December 2004.

[75] Martin Clayton. *Time in Indian Music: Rhythm, Metre, and Form in North Indian Rag Performance.* Oxford Monographs on Music. Oxford University Press, New York, August 2008.

[76] Nick Collins. A comparison of sound onset detection algorithms with emphasis on psychoacoustically motivated detection functions. In *Audio Engineering Society Convention 118*, May 2005.

[77] Nick Collins. Towards a style–specific basis for computational beat tracking. In *9th International Conference on Music Perception and Cognition (ICMPC)*, pages 461–467, Bologna, Italy, Aug. 22-26 2006.

[78] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley-Interscience, 2nd Edition, New Jersey, USA, July 2006.

[79] Michael Scott Cuthbert and Christopher Ariza. Music21: A toolkit for computer-aided musicology and symbolic music data. In *Proc. of the 11th International Society for Music Information Retrieval Conference*, pages 637–642, Utrecht, The Netherlands, August 9-13 2010.

[80] M. Davies and S. Böck. Evaluating the evaluation measures for beat tracking. In *Proc. of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, pages 637–642, Taipei, Taiwan, Oct. 2014.

[81] M. Davies, N. Degara, and M. Plumbley. Evaluation methods for musical audio beat tracking algorithms. Technical Report C4DM-TR-09-06, Center for Digital Music, Queen Mary University, London, UK, Oct. 2009.

[82] Matthew Davies, Guy Madison, Pedro Silva, and Fabien Gouyon. The effect of microtiming deviations on the perception of groove in short rhythms. *Music Perception: An Interdisciplinary Journal*, 30(5):497–510, 2013.

[83] Guilherme de Alencar Pinto. *Razones locas: el paso de Eduardo Mateo por la música uruguaya.* Metro, Montevideo, 1994.

[84] Paulo de Carvalho Neto. The candombe: a dramatic dance from Afro-Uruguayan folklore. *Ehtnomusicology*, VI(3), 1962.

[85] Paulo de Carvalho Neto. *El negro uruguayo: hasta la abolición.* Editorial Universitaria, Quito, 1965.

[86] Centro de Documentación Musical Lauro Ayestarán. Ministerio de Educación y Cultura Uruguay. *Documentos del archivo Ayestarán. La llamada de tambores afromontevideanos entre 1949 y 1966*, 2012. CDM/CD 01.

[87] Centro de Documentación Musical Lauro Ayestarán. Ministerio de Educación y Cultura Uruguay. *Músicas tradicionales del Uruguay. Los toques de los tambores afromontevideanos*, 2012. CDM/DVD 01.

[88] Centro de Documentación Musical Lauro Ayestarán. Ministerio de Educación y Cultura Uruguay. *Músicas tradicionales del Uruguay. Testimonios de tambores afromontevideanos*, 2015. CDM/DVD 02.

[89] Isidoro de María and Vicente Rossi. Cantos y bailes negros. In *Enciclopedia Uruguaya*. Editorial Arca, Montevideo, Uruguay, 1968.

[90] S. de Rooij and P. Vitányi. Approximating Rate-Distortion Graphs of Individual Data: Experiments in Lossy Compression and Denoising. *IEEE Transactions on Computers*, 61(3):395–407, March 2012.

[91] N. Degara, E.A. Rua, A. Pena, S. Torres-Guijarro, M.E.P. Davies, and M.D. Plumbley. Reliability-informed beat tracking of musical signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1):290–301, January 2012.

[92] Christian Dittmar, Martin Pfleiderer, and Meinard Müller. Automated estimation of ride cymbal swing ratios in jazz recordings. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, Málaga, Spain, October 2015.

[93] S. Dixon. Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30(1):39–58, Aug. 2001.

[94] Simon Dixon. Onset detection revisited. In *Proc. of the 9th International Conference on Digital Audio Effects*, pages 133–137, Montreal, Canada, Sep. 2006.

[95] Simon Dixon, Fabien Gouyon, and Gerhard Widmer. Towards characterisation of music via rhythmic patterns. In *Proc. of the 5th International Conference on Music Information Retrieval*, Barcelona, Spain, October 10-14 2004.

[96] S. Durand, J. P. Bello, B. David, and G. Richard. Robust downbeat tracking using an ensemble of convolutional networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(1):76–89, Jan 2017.

[97] John Eargle. *The Microphone Book*. Focal Press, Oxford, 2nd edition, 2004.

[98] Tuomas Eerola. Expectancy-Violation and Information-Theoretic Models of Melodic Complexity. *Empirical Musicology Review*, 11(1), 2016.

[99] D. P. W. Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, May 2007.

[100] Florian Eyben, Sebastian Böck, Björn W. Schuller, and Alex Graves. Universal onset detection with bidirectional long short-term memory neural networks. In *Proc. of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 589–594, Utrecht, Netherlands, Aug. 9-13 2010.

[101] Hugo Fastl and Eberhard Zwicker. *Psychoacoustics: Facts and Models (Springer Series in Information Sciences)*. Springer, Berlin, Germany, third edition, 2006.

[102] Tom Fawcett. ROC Graphs: Notes and Practical Considerations for researchers. Tech Report HPL-2003-4, improved version, HP Labs, 1985.

[103] Patricia Fernández Rielli. Resignificando Cultura e Identidad Afrouruguaya: Interculturalidad en el Candombe. *Revista Trama*, 7(7):108–116, Dec. 2016.

[104] Luis Ferreira. El repicado del Candombe. In *V Jornadas Argentinas de Musicología, Instituto Nacional de Musicología "Carlos Vega", Bs.As.*, 1990.

[105] Luis Ferreira. La música de las llamadas de tambores afrouruguayas: una aproximación a sus características estructurales formales. In Irma Ruiz and Miguel A. García, editors, *Texto y contexto en la investigación musicológica*. Instituto Nacional de Musicología Carlos Vega, Buenos Aires, 1995.

[106] Luis Ferreira. *Los tambores del Candombe*. Ediciones Colihue-Sepé, Montevideo, Uruguay, 1997.

[107] Luis Ferreira. Las Llamadas de Tambores: Comunidad e Identidad de los Afrouruguayos. Master's thesis, Mestrado em Antropologia Social, Universidade de Brasilia, Brasilia, Brazil, 1999.

[108] Luis Ferreira. La música afrouruguaya de tambores en la perspectiva cultural afro-atlántica. In Sonnia Romero Gorski, editor, *Anuario Antropología Social y Cultural en Uruguay*, number 2. Departamento de Antropología Social, Facultad de Humanidades y Ciencias de la Educación (Universidad de la República), Ediciones Nordan Comunidad, Montevideo, 2001.

[109] Luis Ferreira. *El movimiento negro en Uruguay (1988–1998): una versión posible*. Ediciones Étnicas: Mundo Afro, Montevideo, Uruguay, 2003.

[110] Luis Ferreira. An Afrocentric Approach to Musical Performance in the Black South Atlantic: The Candombe Drumming in Uruguay. *TRANS-Transcultural Music Review*, 11:12, 2007.

[111] Luis Ferreira. Dimensiones afro-céntricas en la cultura performática uruguaya. In Gustavo Goldman, editor, *Cultura y sociedad afro-rioplatense*, pages 91–123. Ediciones Perro Andaluz, Montevideo, Uruguay, 2008.

[112] Luis Ferreira. Concepciones cíclicas y rugosidades del tiempo en la práctica musical afroamericana. In Coriún Aharonián, editor, *La música entre África y América*, pages 231–261. Centro Nacional de Documentación Musical Lauro Ayestarán, Montevideo, Uruguay, 2013.

[113] Derry Fitzgerald and Jouni Paulus. Unpitched Percussion Transcription. In Anssi Klapuri and Manuel Davy, editors, *Signal Processing Methods for Music Transcription*, chapter 5, pages 131–162. Springer, New York, USA, 2006.

[114] Neville H. Fletcher and Thomas D. Rossing. *The Physics of Musical Instruments.* Springer, New York, USA, 2nd edition, 2010.

[115] Michel Foucault. *Genealogía del racismo.* Editorial Altamira, La Plata, Argentina, 1996.

[116] Ana Frega, Karla Chagas, Óscar Montaño, and Natalia Stalla. Breve historia de los afrodescendientes en el Uruguay. In Lucía Scuro Somma, editor, *Población afrodescendiente y desigualdades étnico-raciales en Uruguay*, pages 5–102. Programa de las Naciones Unidas para el Desarrollo (PNUD), Montevideo, Uruguay, 2008.

[117] H.G. Gauch. *Scientific Method in Practice.* Cambridge University Press, 2003.

[118] Gery Geenens. Probit transformation for kernel density estimation on the unit interval. *Journal of the American Statistical Association*, 109(505):346–358, 2014.

[119] Christiane Gerischer. O suingue baiano: Rhythmic feeling and microrhythmic phenomena in Brazilian percussion. *Ethnomusicology*, 50(1):99–119, 2006.

[120] Allen Gersho and Robert M. Gray. *Vector Quantization and Signal Compression.* Number 159 in The Kluwer International Series in Engineering and Computer Science. Kluwer, 1992.

[121] Olivier Gillet and Gaël Richard. Enst-drums: an extensive audio-visual database for drum signals processing. In *Proc. of the 7th International Conference on Music Information Retrieval (ISMIR 2006)*, pages 156–159, Victoria, Canada, Oct. 2006.

[122] Olivier Gillet and Gaël Richard. Transcription and separation of drum signals from polyphonic music. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(3):529–540, March 2008.

[123] Paul Gilroy. *The Black Atlantic: Modernity and Double Consciousness.* Verso, London, New York, 1993.

[124] Gustavo Goldman. Seguí que te están mirando. Una primera aproximación al diálogo de los tambores montevideanos. In *VI Jornadas Argentinas de Musicología, Instituto Nacional de Musicología "Carlos Vega", Bs.As.*, 1991.

[125] Gustavo Goldman. Le candombe afro-montevidean. *Musiques Traditionnelles*, (151), 1997.

[126] Gustavo Goldman. *¡Salve Baltasar! La fiesta de reyes en el barrio sur de Montevideo.* re-edited by Perro Andaluz, 2004, Montevideo, 1997.

[127] Gustavo Goldman. *Lucamba. Herencia africana en el tango. 1870-1890.* Ediciones Perro Andaluz, Montevideo, 2008.

[128] Gustavo Goldman. Tango: emergentes de un conflicto en la sociedad afro-montevideana (1867–1890). In Gustavo Goldman, editor, *Cultura y sociedad afro-rioplatense*, pages 175–202. Ediciones Perro Andaluz, Montevideo, Uruguay, 2008.

[129] Alejandro Gortázar. *El licenciado negro: Jacinto Ventura de Molina.* Ediciones Trilce, Montevideo, Uruguay, 2007.

[130] Alejandro Gortázar, José Manuel Barrios, and Adriana Pitetta. *Jacinto Ventura de Molina: Antología de manuscritos (1817-1837).* Universidad de la República, Facultad de Humanidades y Ciencias de la Educación, Montevideo, Uruguay, 2008.

[131] Masataka Goto. AIST annotation for the RWC music database. In *Proc. of the 7th International Conference on Music Information Retrieval (ISMIR 2006)*, pages 359–360, Victoria, Canada, Oct. 2006.

[132] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka. RWC music database: Popular, classical and jazz music databases. In *Proc. of the 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, Paris, France, Oct. 2002.

[133] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka. RWC music database: Music genre database and musical instrument sound database. In *Proc. of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, Baltimore, Maryland, USA, Oct. 2003.

[134] Peter D. Grünwald, In Jae Myung, and Mark A. Pitt. *Advances in Minimum Description Length: Theory and Applications (Neural Information Processing).* The MIT Press, 2005.

[135] P. Grünwald and P. Vitany. Shannon Information and Kolmogorov Complexity. *arXiv preprint*, cs/040002, 2004, updated 2010.

[136] Gérald Guillot. *From implicit musical objects to their exogenous formal didactising: internal didactic transposition of suíngue brasileiro in France.* PhD thesis, Paris-Sorbonne University, Paris, France, May 2011.

[137] Emilia Gómez, Perfecto Herrera, and Francisco Gómez-Martin. Computational ethnomusicology: perspectives and challenges. *Journal of New Music Research*, 42(2):111–112, 2013.

[138] Martín Haro and Perfecto Herrera. From low-level to song-level percussion descriptors of polyphonic music. In *Proceedings of the 10th International Society for Music Information Retrieval Conference, (ISMIR 2009)*, pages 243–248, Kobe, Japan, October 26-30 2009. International Society for Music Information Retrieval.

[139] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[140] Perfecto Herrera, Amaury Dehamel, and Fabien Gouyon. Automatic labeling of unpitched percussion sounds. In *AES 114th Convention*, pages 1–14, Amsterdam, The Netherlands, March 2003. Convention Paper 5806.

[141] A. Holzapfel, F. Krebs, and A. Srinivasamurthy. Tracking the "odd": Meter inference in a culturally diverse music corpus. In *Proc. of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, pages 425–430, Taipei, Taiwan, Oct. 2014.

[142] Andre Holzapfel. Relation between surface rhythm and rhythmic modes in Turkish makam music. *Journal for New Music Research*, 44(1):25–38, 2015.

[143] Andre Holzapfel and Thomas Grill. Bayesian meter tracking on learned signal representations. In *Proceedings of ISMIR - International Conference on Music Information Retrieval*, pages 262–268, 2016.

[144] Henkjan Honing. From time to time: The representation of timing and tempo. *Computer Music Journal*, 25(3):50–61, August 2001.

[145] Henkjan Honing. The final ritard: On music, motion, and kinematic models. *Computer Music Journal*, 27(3):66–72, August 2003.

[146] David Huron. Music information processing using the humdrum toolkit: Concepts, examples, and lessons. *Computer Music Journal*, 26(2):11–26, July 2002.

[147] Jason L. Hutchens and Michael D. Alder. Finding structure via compression. In *Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning*, NeMLaP3/CoNLL '98, pages 79–82, Stroudsburg, PA, USA, 1998. Association for Computational Linguistics.

[148] Ema Isola. *La esclavitud en el Uruguay: desde sus comienzos hasta su extinción (1743–1852)*. Comisión nacional de homenaje del sesquicentenario de los hechos históricos de 1825, Montevideo, 1975.

[149] Vijay Iyer. Embodied mind, situated cognition, and expressive microtiming in African-American music. *Music Perception: An Interdisciplinary Journal*, 19(3):387–414, 2002.

[150] Anil K. Jain. Data Clustering: 50 Years Beyond K-Means. *Pattern Recognition Letters*, 31(8):651–666, June 2010.

[151] Berit Janssen, W. Bas de Haas, Anja Volk, and Peter van Kranenburg. Finding repeated patterns in music: State of knowledge, challenges, perspectives. In *Proceedings of the 10th International Symposium on Computer*

Music Multidisciplinary Research (CMMR 2013), pages 225–240, Marseille, France, Oct. 15-18 2013.

[152] Tristan Jehan. *Creating Music by Listening.* PhD thesis, MIT, Cambridge, USA, September 2005.

[153] Arthur M. Jones. African rhythm. *Africa: Journal of the International African Institute*, 24(1):26–47, 1954.

[154] Arthur M. Jones. *Studies in African Music.* Oxford University Press, London, 1959.

[155] Patrick Juola. Assessing linguistic complexity. In Kaius Sinnemaki Miestamo, Matti and Fred Karlsson, editors, *Language Complexity: Typology, Contact, Change*, chapter 6, pages 89–108. John Benjamins Press, Amsterdam, 2007.

[156] Luis Jure, , and Olga Picún. Los cortes de los tambores. Aspectos musicales y funcionales de las paradas en las llamadas de tambores afro montevideanos. In *VII Jornadas Argentinas de Musicología, Instituto Nacional de Musicología "Carlos Vega", Bs.As.*, 1992.

[157] Luis Jure. ¡Perico, suba ahí! Pautación y análisis de un solo de repique de Pedro 'Perico' Gularte. In *VII Jornadas Argentinas de Musicología– VI Conferencia Anual de la Asociación Argentina de Musicología*, Córdoba, Argentina, October 1992. In Spanish.

[158] Luis Jure. Principios generativos del toque de repique del candombe. In Coriún Aharonián, editor, *La música entre África y América*, pages 263–291. Centro Nacional de Documentación Musical Lauro Ayestarán, Montevideo, Uruguay, 2013.

[159] Luis Jure and Martín Rocamora. Microtiming in the rhythmic structure of candombe drumming patterns. In *Fourth International Conference on Analytical Approaches to World Music*, New York, USA, June 8-11 2016.

[160] Luis Jure and Martín Rocamora. Clave patterns in Uruguayan candombe drumming. In *16th Rhythm Production and Perception Workshop*, Birmingham, UK, July 3-5 2017.

[161] Rebecka Jörnsten. Simultaneous model selection via rate-distortion theory, with applications to cluster and significance analysis of gene expression data. *Journal of Computational and Graphical Statistics*, 18(3):613–639, 2009.

[162] L. Kaufman and Peter J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis.* John Wiley, New Jersey, USA, 1990.

[163] Michael J. Kearns and Umesh V. Vazirani. *An Introduction to Computational Learning Theory.* MIT Press, Cambridge, MA, USA, 1994.

[164] Arik Kershenbaum. Entropy rate as a measure of animal vocal complexity. *Bioacoustics*, 23(3):195–208, 2014.

[165] Elizabeth W. Kiddy. Congados, Calunga, Candombe: Our Lady of the Rosary in Minas Gerais, Brazil. *Luso-Brazilian Review*, 37(1):47–61, 2000.

[166] A. P. Klapuri, A. J. Eronen, and J. T. Astola. Analysis of the meter of acoustic musical signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):342–355, Jan. 2006.

[167] Anssi Klapuri. Sound onset detection by applying psychoacoustic knowledge. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, volume 6, pages 3089–3092, Mar 1999.

[168] Anssi Klapuri. Pattern induction and matching in music signals. In *Proceedings of the 7th International Symposium on Computer Music Modeling and Retrieval*, CMMR'10, pages 188–204, Berlin, Heidelberg, 2011. Springer-Verlag.

[169] Mieczyslaw Kolinski. A cross-cultural approach to metro-rhythmic patterns. *Ethnomusicology*, 17(3):494–506, 1973.

[170] Andrey N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems in Information Transmission*, 1:1–7, 1965.

[171] Ioannis Kontoyiannis. Model selection via rate-distortion theory. In *Proc. of the 34th Annual Conference on Information Sciences and Systems*, pages 1–5, Princeton, NJ, Mar., 15-17 2000.

[172] F. Krebs, F. Korzeniowski, M. Grachten, and G. Widmer. Unsupervised learning and refinement of rhythmic patterns for beat and downbeat tracking. In *22nd European Signal Processing Conference (EUSIPCO)*, pages 611–615, Sept 2014.

[173] Florian Krebs, Sebastian Böck, Matthias Dorfer, and Gerhard Widmer. Downbeat estimation using beat synchronous features and recurrent neural networks. In *Proc. of the 17th International Society for Music Information Retrieval Conference*, pages 129–135, New York, USA, August 2016.

[174] Florian Krebs, Sebastian Böck, and Gerhard Widmer. Rhythmic pattern modeling for beat and downbeat tracking in musical audio. In *Proc. of the 14th International Society for Music Information Retrieval Conference*, pages 227–232, Curitiba, Brazil, November 2013.

[175] Florian Krebs, Sebastian Böck, and Gerhard Widmer. An efficient state-space model for joint tempo and meter tracking. In *16th International Society for Music Information Retrieval (ISMIR) Conference*, pages 72–78, Malaga, Spain, October 2015.

[176] Florian Krebs, Andre Holzapfel, Ali Taylan Cemgil, and Gerhard Widmer. Inferring metrical structure in music using particle filters. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, 23(5):817–827, May 2015.

[177] Gerhard Kubik. Drum patterns in the "Batuque" of Benedito Caxias. *Latin American Music Review / Revista de Música Latinoamericana*, 11(2):115–181, 1990.

[178] Jill Lane. *Blackface Cuba, 1840-1895.* University of Pennsylvania Press, Philadelphia, 2005.

[179] Marc Leman. Systematic musicology at the crossroads of modern music research. In A. Schneider, editor, *Systematic and comparative musicology: concepts, methods, findings*, pages 89–115. Karl Dieter Wagner., Hamburger Jahrbuch für Musikwissenschaft, 24. Frankfurt am Main: Peter Lang, 2008.

[180] A. Lempel and J. Ziv. On the complexity of finite sequences. *IEEE Transactions on Information Theory*, 22(1):75–81, Jan 1976.

[181] F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music.* MIT Press, Cambridge, USA, 1985.

[182] Micheline Lesaffre, Marc Leman, Bernard De Baets, and Jean-Pierre Martens. Methodological considerations concerning manual annotation of musical audio in function of algorithm development. In *Proc. of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*, Barcelona, Spain, Oct. 2004.

[183] S. Leung-Yan-Cheong and T. Cover. Some equivalences between Shannon entropy and Kolmogorov complexity. *IEEE Transactions on Information Theory*, 24(3):331–338, May 1978.

[184] D.J. Levitin. *This is Your Brain on Music: The Science of a Human Obsession.* Penguin Group, New York, USA, 2006.

[185] Marvin A. Lewis. *Afro-Uruguayan Literature: Postcolonial Perspectives.* Associated University Press, London, 2003.

[186] Ming Li, Xin Chen, Xin Li, Bin Ma, and P. M. B. Vitányi. The similarity metric. *IEEE Transactions on Information Theory*, 50(12):3250–3264, Dec 2004.

[187] Ming Li and Paul M. B. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications.* Springer Publishing Company, 3 edition, 2008.

[188] Yoseph Linde, Andrés Buzo, and Robert M. Gray. An algorithm for vector quantizer design. *IEEE Transactions on Communications*, 28(1):84–95, Jan 1980.

[189] Stuart P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, Mar 1982.

[190] David Locke. Principles of offbeat timing and cross-rhythm in Southern Eve dance drumming. *Ethnomusicology*, 26(2):217–246, 1982.

[191] A. Londei, V. Loreto, and M. O. BelardinellKirill. Musical style and authorship categorization by informative compressors. In *Proc. of the 5th Triennial ESCOM Conference*, pages 200–203, Hanover, Germany, Sep. 2003.

[192] J. London. Cognitive constraints on metric systems: Some observations and hypotheses. *Music Perception*, 19(4):529—-550, Summer 2002.

[193] J. London. *Hearing in Time: Psychological Aspects of Musical Meter*. Oxford University Press, New York, USA, 2004.

[194] Justin London, Rainer Polak, and Nori Jacoby. Rhythm histograms and musical meter: A corpus study of Malian percussion music. *Psychonomic Bulletin & Review*, 24(2):474–480, Apr 2017.

[195] Eric Lott. Love and theft: The racial unconscious of blackface minstrelsy. *Representations*, Summer(39):23–50, 1992.

[196] Corentin Louboutin and David Meredith. Using general-purpose compression algorithms for music analysis. *Journal of New Music Research*, 45(1):1–16, 2016.

[197] Bernardo Marenco, Magdalena Fuentes, Florencia Lanzaro, Martín Rocamora, and Alvaro Gómez. A multimodal approach for percussion music transcription from audio and video. In *Proceedings of the 20th Iberoamerican Congress on Pattern Recognition (CIARP 2015). Montevideo, Uruguay*, 9-12 nov. 2015. link.springer.com.

[198] M. M. Marin and H. Leder. Examining complexity across domains: Relating subjective and objective measures of affective environmental scenes, paintings and music. *PLoS ONE*, 8(8):e72412, 2013.

[199] Paul Masri. *Computer Modeling of Sound for Transformation and Synthesis of Musical Signals*. PhD thesis, University of Bristol, UK, 1996.

[200] Matthias Mauch and Simon Dixon. A corpus-based study of rhythm patterns. In *Proc. of the 13th International Society for Music Information Retrieval Conference*, Porto, Portugal, October 8-12 2012.

[201] Panayotis Mavromatis. Minimum description length modelling of musical structure. *Journal of Mathematics and Music*, 3(3):117–136, 2009.

[202] M. F. McKinney and D. Moelants. Deviations from the resonance theory of tempo induction. In *Proc. of the Conference on Interdisciplinary Musicology (CIM04)*, pages 1–11, Graz, Austria, Apr. 2004.

[203] David Meredith. Music analysis and kolmogorov complexity. In *Proceedings of the 19th Colloquio di Informatica Musicale (XIX CIM)*, pages 1–7, Trieste, Italy, Nov. 21-24 2012.

[204] David Meredith. Music analysis and point-set compression. *Journal of New Music Research*, 44(3):245–270, 2015.

[205] David Meredith. Analysing music with point-set compression algorithms. In David Meredith, editor, *Computational Music Analysis*, pages 335–366. Springer International Publishing, 2016.

[206] Francisco M. Merino. *El negro en la sociedad montevideana*. Banda Oriental, Montevideo, 1982.

[207] Hristina Mikić. Measuring the economic contribution of cultural industries: A review and assessment of current methodological approaches. In UNESCO, editor, *Framework for cultural statistics handbook*. UNESCO Institute for Statistics, Montreal, Canada, 2012.

[208] Marius Miron, Matthew E.P. Davies, and Fabien Gouyon. An open-source drum transcription system for Pure Data and Max MSP. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 221–225, May 2013.

[209] Dirk Moelants. The performance of notes inégales: The influence of tempo, musical structure, and individual performance style on expressive timing. *Music Perception: An Interdisciplinary Journal*, 28(5):449–460, 2011.

[210] Musique Du Monde. *Uruguay: Tambores del Candombe*, 2000. Buda Records, 92745-2.

[211] Óscar Montaño. *Umkhonto: Historia del aporte negro-africano en la formacion del Uruguay*. Rosebud Ediciones, Montevideo, Uruguay, 1997.

[212] Óscar Montaño. *Historia afrouruguaya, Volume 1*. Mastergraf, Montevideo, Uruguay, 2008.

[213] M. M uller, D. P. W. Ellis, A. Klapuri, and G. Richard. Signal processing for music analysis. *IEEE Journal of Selected Topics in Signal Processing*, 5(6):1088–1110, Oct 2011.

[214] K. P. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, Computer Science Division, University of California, Berkeley, USA, Fall 2002.

[215] Meinard Müller. Music structure analysis. Chapter 4. In *Fundamentals of Music Processing*, pages 167–236. Springer Verlag, 2015.

[216] Luiz Naveda, Fabien Gouyon, Carlos Guedes, and Marc Leman. Microtiming patterns and interactions with musical properties in samba music. *Journal of New Music Research*, 40(3):225–238, 2011.

[217] Kwabena J.H. Nketia. *The Music of Africa*. W. W. Norton, New York, 1974.

[218] Leonardo Nunes. *Algoritmos para análise rítmica computacional.* PhD Thesis, Universidade Federal do Rio de Janeiro. Rio de Janeiro, Brazil, 2014.

[219] Leonardo Nunes, Martín Rocamora, Luis Jure, and Luiz W. P. Biscainho. Beat and downbeat tracking based on rhythmic patterns applied to the Uruguayan candombe drumming. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015). Málaga, Spain*, pages 264–270, 26-30 oct. 2015. ismir2015.uma.es.

[220] J. L. Oliveira, M. E. P. Davies, F. Gouyon, and L. P. Reis. Beat tracking for multiple applications: A multi-agent system architecture with state recovery. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(10):2696–2706, Dec. 2012.

[221] Rubén Olivera. Introducción a la música popular uruguaya 1973-2013. In *Música*, volume 5 of *Libro de los Bicentenarios*, pages 5–50. Comisión del Bicentenario, IMPO, Montevideo, Uruguay, 2013.

[222] Tomás Olivera Chirimini. Candombe, African Nations, and the Africanity of Uruguay. In Sheila S. Walker, editor, *African Roots/American Cultures: Africa in the Creation of the Americas*, pages 256–74. Rowman & Littlefield Publishers, New York, 2001.

[223] Tomás Olivera Chirimini and Juan Antonio Varese. *Memorias del tamboril.* Latina, re-edited by Banda Oriental (2002), Montevideo, 1996.

[224] Tomás Olivera Chirimini and Juan Antonio Varese. *Los Candombes de Reyes. Las Llamadas.* El Galeón, Montevideo, 2000.

[225] Antonio Ortega and Kannan Ramchandran. Rate-distortion methods for image and video compression. *IEEE Signal Processing Magazine*, 15(6):23–50, Nov 1998.

[226] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, January 1979.

[227] Roberto Pacheco. Invisible but not forgotten: The Afro-Argentine and Afro-Uruguayan experience from the sixteenth to the twentieth centuries. Master's thesis, Florida International University, Miami, Florida, 2001.

[228] Roberto Pacheco. Bibliografía afro-rioplatense (1999-2003): invisible, pero no olvidada. In Gustavo Goldman, editor, *Cultura y sociedad afro-rioplatense*, pages 11–45. Ediciones Perro Andaluz, Montevideo, Uruguay, 2008.

[229] Roberto Pacheco. Estudios sobre negros en Argentina y Uruguay en el nuevo milenio: bibliografía afro-rioplatense (1999-2003). In Gustavo Goldman, editor, *Cultura y sociedad afro-rioplatense*, pages 49–58. Ediciones Perro Andaluz, Montevideo, Uruguay, 2008.

[230] Jouni K. Paulus and Anssi P. Klapuri. Conventional and periodic n-grams in the transcription of drum sequences. In *Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on*, volume 2, pages II–737–40 vol.2, July 2003.

[231] Jouni K. Paulus and Anssi P. Klapuri. Drum sound detection in polyphonic music with hidden markov models. *Eurasip Journal on Audio, Speech, and Music Processing*, 2009:4, 2009.

[232] G. Peeters and H. Papadopoulos. Simultaneous beat and downbeat-tracking using a probabilistic framework: Theory and large-scale evaluation. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(6):1754–1769, Aug 2011.

[233] Geoffroy Peeters. Template-based estimation of time-varying tempo. *EURASIP Journal on Advances in Signal Processing*, 2007(67215):1–14, Dec. 2006.

[234] Eugenio Petit Muñoz, Edmundo M. Narancio, and José M. Traibel Nelcis. *La condición jurídica, social, económica y política de los negros durante el coloniaje en la Banda Oriental.* Facultad de Derecho y Ciencias Sociales, Montevideo, 1947.

[235] Olga Picún. El candombe y la música popular uruguaya. *Perspectiva Interdisciplinaria de Música, Universidad Nacional Autónoma de México*, 1(1):73–78, Sep. 2006.

[236] Olga Picún. Procesos de resignificación y legitimación del candombe: coincidencias y consecuencias. In Coriún Aharonián, editor, *La música entre África y América*, pages 293–309. Centro Nacional de Documentación Musical Lauro Ayestarán, Montevideo, Uruguay, 2013.

[237] Olga Picún. Apropiaciones, identidades, transformaciones y tensiones. Candombe de hoy. In *Patrimonio vivo del Uruguay. Relevamiento de Candombe*, pages 185–264. Dirección de Cooperación y Proyectos Internacionales y Comisión Nacional del Uruguay para la UNESCO. Ministerio de Educación y Cultura, Montevideo, Uruguay, 2015.

[238] Rainer Polak. Rhythmic feel as meter: Non-isochronous beat subdivision in Jembe music from Mali. *Music Theory Online*, 16(4), 2010.

[239] Rainer Polak and Justin London. Timing and Meter in Mande Drumming from Mali. *Music Theory Online*, 20(1):1–22, March 2014.

[240] Rainer Polak, Justin London, and Nori Jacoby. Both isochronous and non-isochronous metrical subdivision afford precise and stable ensemble entrainment: A corpus study of malian jembe drumming. *Frontiers in Neuroscience*, 10:285, 2016.

[241] Rafael Porzecanski. *El Uruguay judío. Demografía e identidad.* Editorial Trilce, Montevideo, Uruguay, 2006.

[242] Teresa Porzecanski and Beatriz Santos. *Historias de exclusión: afrodescendientes en el Uruguay.* Editorial Linardi y Risso, Montevideo, Uruguay, 2006.

[243] Aníbal Quijano. Coloniality of Power, Eurocentrism, and Latin America. In *Nepantla: Views from South*, number 3, pages 533–580. Duke University Press, Duke-Durham, NC, USA, 2000.

[244] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, Feb 1989.

[245] Rubén Rada and Fernando Peláez. *Rada.* Editorial Aguilar, Montevideo, 2013.

[246] Carlos M. Rama. *Los afro-uruguayos.* El Siglo Ilustrado, Montevideo, 1966.

[247] Tania Ramírez. *Ciudadanía afrodescendiente.* Hablando de derechos. DESC+A. Pensando en derechos humanos. Dirección Nacional de Política Social, Ministerio de Desarrollo Social, Montevideo, Uruguay, 2012.

[248] J. Rigau, M. Feixas, and M. Sbert. An information-theoretic framework for image complexity. In *Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging*, Computational Aesthetics'05, pages 177–184, Aire-la-Ville, Switzerland, Switzerland, 2005. Eurographics Association.

[249] J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, September 1978.

[250] Martín Rocamora. Automatic analysis of percussion music: the Afro-Uruguayan candombe drumming as a case study, October 12–15 2014. Short oral presentation at the *II International Workshop on Cross-disciplinary and Multi-cultural Perspectives on Musical Rhythm and Improvisation*, New York University Abu Dhabi. Abu Dhabi, United Arab Emirates.

[251] Martín Rocamora. Interpersonal music entrainment in Afro-Uruguayan candombe drumming, July 13–19 2017. Oral presentation at the *44th International Council for Traditional Music (ICTM) World Conference*. Limerick, Irleand. Work in collaboration with Nori Jacobi, Rainer Polak and Luis Jure.

[252] Martín Rocamora and Luiz W. P. Biscainho. Modeling onset spectral features for discrimination of drum sounds. In *Proceedings of the 20th Iberoamerican Congress on Pattern Recognition (CIARP 2015). Montevideo, Uruguay*, pages 100–107, 9-12 nov. 2015. link.springer.com.

[253] Martín Rocamora and Luis Jure. Rhythmic pattern tracking: the Uruguayan candombe drumming as a case study, September 26–28 2013. Oral presentation at the *I Congreso Internacional de Ciencia y Tecnología Musical (CICTeM 2013)*, IUNA. Buenos Aires, Argentina.

[254] Martín Rocamora, Luis Jure, and Luiz W. P. Biscainho. Tools for detection and classification of piano drum patterns from candombe recordings. In *Proceedings of the 9th Conference on Interdisciplinary Musicology (CIM 2014)*, pages 382–387, Berlin, Germany, 4-6 dec. 2014. iie.fing.edu.uy.

[255] Martín Rocamora, Luis Jure, Bernardo Marenco, Magdalena Fuentes, Florencia Lanzaro, and Alvaro Gómez. An audio-visual database of candombe performances for computational musicological studies. In *Proceedings of the II Congreso Internacional de Ciencia y Tecnología Musical (CICTeM 2015)*, pages 17–24, Buenos Aires, Argentina, 26-28 sep. 2015. iie.fing.edu.uy.

[256] Romero J. Rodríguez. *Mbundo Malungo a Mundele. Historia del Movimiento Afrouruguayo y sus Alternativas de Desarrollo.* Rosebud Ediciones, Montevideo, Uruguay, 2006.

[257] Miguel Romano. *Toca candombe.* Ediciones Ayuí, Montevideo, Uruguay, 2001.

[258] Carlos Rosão, Ricardo Ribeiro, and David Martins de Matos. Influence of peak selection methods on onset detection. In *Proceedings of the 13th International Society for Music Information Retrieval Conference, (ISMIR 2012)*, pages 517–522, Porto, Portugal, Oct. 8-12 2012.

[259] Vicente Rossi. *Cosas de negros.* Hachette, 2da edición, Buenos Aires, 1958.

[260] Viviana Ruiz. Relevamiento bibliográfico, documental y testimonial. Pasado y presente en el candombe. In *Patrimonio vivo del Uruguay. Relevamiento de Candombe*, pages 7–64. Dirección de Cooperación y Proyectos Internacionales y Comisión Nacional del Uruguay para la UNESCO. Ministerio de Educación y Cultura, Montevideo, Uruguay, 2015.

[261] L.K. Saul, K.Q. Weinberger, J. H. Ham, F. Sha, and D. D. Lee. Spectral methods for dimensionality reduction. In O. Chapelle, B. Schölkopf, and A. Zien, editors, *Semi-Supervised Learning*, chapter 16, pages 293–308. MIT Press, Cambride, USA, 2006.

[262] Markus Schedl, Emilia Gómez, and Julián Urbano. Music information retrieval: Recent developments and applications. *Foundations and Trends in Information Retrieval*, 8(2-3):127–261, 2014.

[263] Eric D. Scheirer. Tempo and beat analysis of acoustic signals. *Journal of the Acoustic Society of America*, 103(1):588–601, 1998.

[264] Rita L. Segato. Racismo, discrminación y acciones afirmativas: herramientas conceptuales. In *Educar en ciudadanía intercultural*. Pontifica Universidad Católica del Perú, Lima, Perú, 2006.

[265] Rita L. Segato. El color de la cárcel en América Latina. Apuntes sobre la colonialidad de la justicia en un continente en desconstrucción. *Revista Nueva sociedad*, Mar.-Apr.(208), 2007.

[266] 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*. Creative Commons BY-NC-ND 3.0 license, 2013.

[267] Xavier Serra. A multicultural approach in music information research. In *Proc. of the 12th International Society for Music Information Retrieval Conference*, pages 151–156, Miami, USA, October 24-28 2011.

[268] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. In *IRE National Convention Record, Part 4*, pages 142–163. 1959.

[269] Claude Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[270] Hideaki Shimazaki and Shigeru Shinomoto. Kernel bandwidth optimization in spike rate estimation. *Journal of Computational Neuroscience*, 29(1):171–182, Aug 2010.

[271] Kirill A. Sidorov, Andrew Jones, and A. David Marshall. Music analysis as a smallest grammar problem. In *Proc. of the 15th International Society for Music Information Retrieval Conference, ISMIR 2014*, pages 301–306, Taipei, Taiwan, Oct. 2014.

[272] Jukka Sillanpää. Drum stroke recognition. Technical report, Tampere University of Technology, Tampere, Finland, 2000.

[273] Malcolm Slaney. Auditory toolbox. Technical Report 1998-010, Interval Research Corporation, 1994.

[274] R.J. Solomonoff. A formal theory of inductive inference. Part I and II. *Information and Control*, 7(1):1–22, 224–254, 1964.

[275] A. Srinivasamurthy, A. Holzapfel, and X. Serra. In search of automatic rhythm analysis methods for turkish and indian art music. *Journal of New Music Research*, 43:94–114, Mar. 2014.

[276] Ajay Srinivasamurthy, Andre Holzapfel, Ali Taylan Cemgil, and Xavier Serra. Particle filters for efficient meter tracking with dynamic bayesian networks. In *16th International Society for Music Information Retrieval (ISMIR) Conference*, pages 197–203, Malaga, Spain, October 2015.

[277] Dan Stowell and Mark Plumbley. Adaptive whitening for improved real-time audio onset detection. In *Proceedings of the International Computer Music Conference (ICMC)*, 2007.

[278] S. Streich. *Music Complexity: a multi-faceted description of audio content.* PhD thesis, Universitat Pompeu Fabra, 2007.

[279] Koen Tanghe, Sven Degroeve, and Bernard De Baets. An algorithm for detecting and labeling drum events in polyphonic music. In *Proceedings of the first Music Information Retrieval Evaluation eXchange (MIREX)*, 2005.

[280] D. Temperley. *The cognition of basic musical structures.* Mit Press, Cambridge, USA, 2001.

[281] David Temperley. Meter and Grouping in African Music: A View from Music Theory. *Ethnomusicology*, 44(1):65–96, 2000.

[282] Lucas Thompson, Simon Dixon, and Matthias Mauch. Drum transcription via classification of bar-level rhythmic patterns. In *Proceedings of the 15th International Society for Music Information Retrieval Conference, (ISMIR 2014)*, pages 187–192, Taipei, Taiwan, October 27-31 2014.

[283] Adam R. Tindale, Ajay Kapur, George Tzanetakis, Peter Driessen, and Andrew Schloss. A comparison of sensor strategies for capturing percussive gestures. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME '05)*, pages 200–203, Vancouver, Canada, 2005.

[284] Godfried T Toussaint et al. A mathematical analysis of African, Brazilian, and Cuban clave rhythms. In *Proceedings of BRIDGES: Mathematical Connections in Art, Music and Science*, pages 157–168, 2002.

[285] Godfried T Toussaint et al. Classification and phylogenetic analysis of African ternary rhythm timelines. In *Proceedings of BRIDGES: Mathematical Connections in Art, Music and Science*, pages 25–36, 2003.

[286] G. Tzanetakis, A. Kapur, W. A. Schloss, and M. Wright. Computational ethnomusicology. *Journal of interdisciplinary music studies*, 1(2):1–24, 2007.

[287] UNESCO. Representative list of the intangible cultural heritage of humanity – Candombe and its socio–cultural space: a community practice, 2009. `http://www.unesco.org/culture/ich/RL/00182`.

[288] Idelfonso Pereda Valdés. *El negro en el Uruguay, pasado y presente*. author's edition, Montevideo, 1965.

[289] Robert Van Rooyen and George Tzanetakis. Pragmatic drum motion capture system. In *Proc. of the International Conference on New Interfaces for Musical Expression (NIME '15)*, Lousiana, USA, 2015.

[290] N. K. Vereshchagin and P. M. B. Vitányi. Rate Distortion and Denoising of Individual Data Using Kolmogorov Complexity. *IEEE Transactions on Information Theory*, 56(7):3438–3454, July 2010.

[291] Anja Volk, Frans Wiering, and Peter van Kranenburg. Unfolding the potential of Computational Musicology. In *Proc. of the International Conference on Informatics and Semiotics in Organisations*, pages 137–144, Leeuwarden, Netherlands, Nov. 2011.

[292] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17:395–416, 2007.

[293] N. Whiteley, A. Cemgil, and S. Godsill. Bayesian modelling of temporal structure in musical audio. In *Proc. of the 7th Int. Conference on Music Information Retrieval*, pages 29–34, Victoria, Canada, October 2006.

[294] N. Whiteley, A. Cemgil, and S. Godsill. Sequential inference of rhythmic structure in musical audio. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1321–1324, Honolulu, USA, April 2007.

[295] M. Wright, W. A. Schloss, and G. Tzanetakis. Analyzing Afro-Cuban rhythms using rotation-aware clave template matching with dynamic programming. In *Proc. of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, pages 647–652, Philadelphia, USA, Sep. 2008.

[296] J. R. Zapata, M. E. P. Davies, and E. Gómez. Multi-feature beat tracking. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(4):816–825, April 2014.

[297] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23(3):337–343, May 1977.

Glossary

bar group of certain number of beats defining the highest metrical level.

beats points in time that define the pulse of the music.

bpm beats per minute, the unit used to measure tempo.

downbeat the first beat of a measure, bar or rhythmic cycle.

measure see bar, bar and measure are used interchangeably.

meter a structure of regular points in time hierarchically organized in levels.

rhythm patterns of organised durations that are present in the music.

tactus pulsation of the perceptually most salient metrical level.

tatum the lowest metrical pulse (subdivision).

tempo the frequency of the beats of the most salient metrical level.

This is the last page.
Compiled on Tuesday 11$^{\text{th}}$ September, 2018.
`http://iie.fing.edu.uy/~rocamora/phdthesis/`