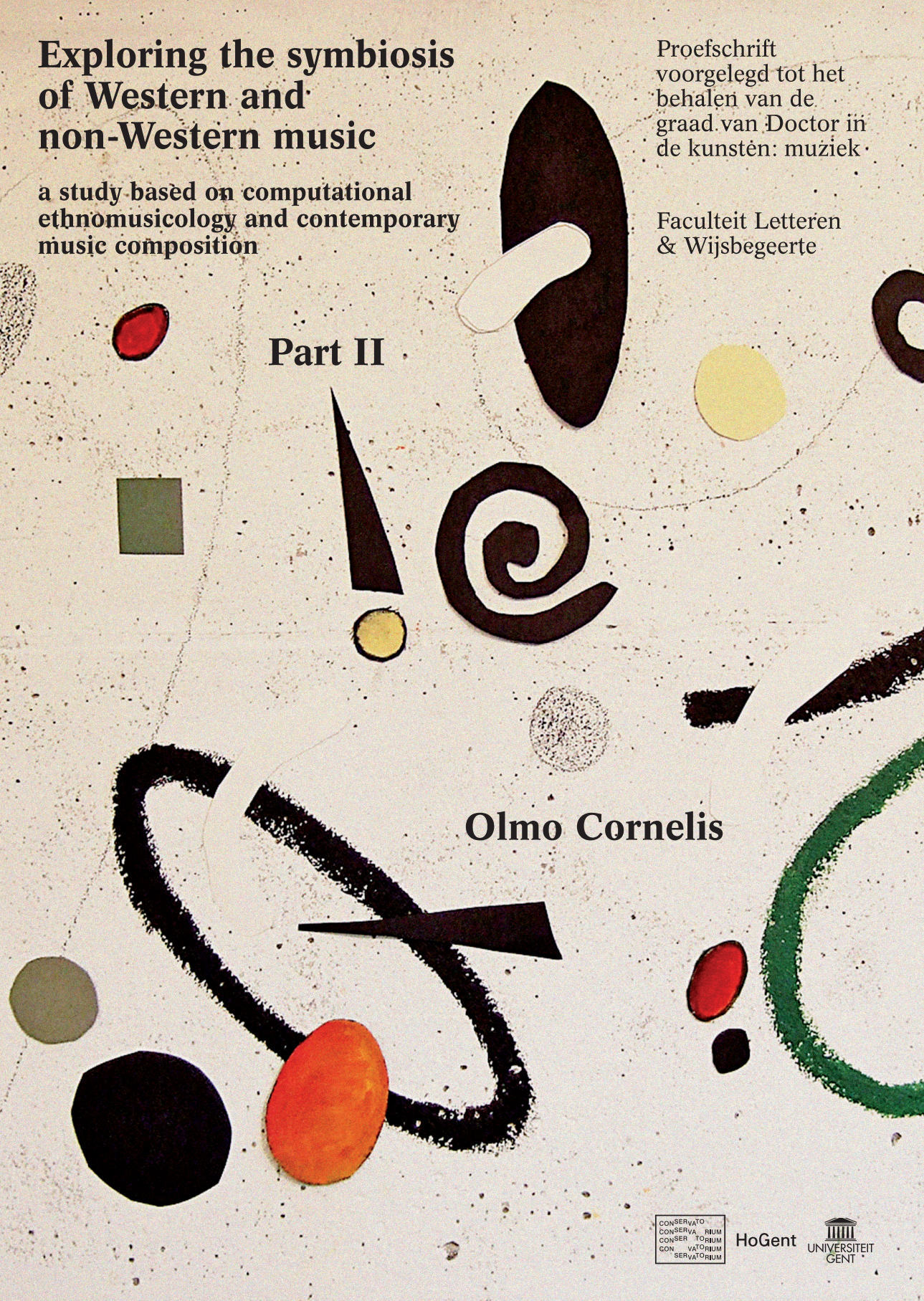
Exploring the symbiosis of Western and non-Western music

a study based on computational ethnomusicology and contemporary music composition

Proefschrift voorgelegd tot het behalen van de graad van Doctor in de kunsten: muziek

Faculteit Letteren & Wijsbegeerte

Part II



Olmo Cornelis

CONSERVATORIUM CONSERVATORIUM CONSERVATORIUM CONSERVATORIUM
HoGent
UNIVERSITEIT GENT

To
Ti Ta Thijmen,
mini Mauro,
and an amazing Anna

Promotoren	Prof. Dr. Marc Leman Vakgroep Kunst-, Muziek- en Theaterwetenschappen Lucien Posman Vakgroep Muziekcreatie, School of Arts, Hogeschool Gent
Decaan	Prof. Dr. Marc Boone
Rector	Prof. Dr. Anne De Paepe

Leescommissie

Dr. Micheline Lesaffre
Prof. Dr. Francis Maes
Dr. Godfried-Willem Raes
Peter Vermeersch
Dr. Frans Wiering

Aanvullende examencommissie

Prof. Dr. Jean Bourgeois (voorzitter)
Prof. Dr. Maximiliaan Martens
Prof. Dr. Dirk Moelants
Prof. Dr. Katharina Pewny
Prof. Dr. Linda Van Santvoort

Kaftinformatie: Art work by Noel Cornelis, cover by Inge Ketelers

ISBN: 978-94-6197-256-9

Olmo Cornelis has been affiliated as an artistic researcher to the Royal Conservatory, School of Arts Ghent since February 2008. His research project was funded by the Research Fund University College Ghent.

UNIVERSITEIT GENT

Faculteit Letteren & Wijsbegeerte

Olmo Cornelis

Exploring the symbiosis of Western and non-Western music

a study based on computational ethnomusicology and contemporary music composition

Part II

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de kunsten: muziek

2013

Introduction

This book is the second book in a series of three that constitutes my doctoral dissertation. This part contains a compilation of the main articles from my research that is described in book I.

All articles are listed in the next section. Items that are displayed in bold are included in this part, and are listed in the table of contents.

Published articles

International peer reviewed publications (A1):

Cornelis, O., Lesaffre, M., Moelants, D., & Leman, M. (2010). Access to ethnic music: Advances and Perspectives in Content based Music Information Retrieval. Signal Processing, 90(4) pp. 1008–1031.

Six, J., Cornelis, O., & Leman, M. (2013). Tarsos, a Modular Platform for precise Pitch Analysis of Western and non-Western music. Journal of New Music Research. 42 (2) pp. 113-129.

Cornelis, O., Six, J., Holzapfel, A., & Leman, M. (2013). Evaluation and Recommendation of Pulse and Tempo Annotation in Ethnic Music. Journal of New Music Research. 42 (2) pp. 131-149.

Cornelis, O. (2013). From Information to Inspiration, Sensitivities mapped in a Casus of Central-African Music Analysis and Contemporary Music Composition. Critical Arts. 27 (5) pp. 624-635.

Lidy T., Silla C.N., Cornelis O., Gouyon F., Rauber A. ,Kaestner C.A.A., Koerich A.L. (2010). 'Western vs. Ethnic Music: On the Suitability of State-of-the-art Music Retrieval Methods for Analyzing, Structuring and Accessing Ethnic Music Collections', Signal Processing, Elsevier. 90 pp. 1032–1048.

Matthé, T., De Caluwe, R., De Tré, G., Hallez, A., Verstraete, J., Leman, M., Cornelis, O., et al. (2006). Similarity between multi-valued thesaurus attributes: theory and application in multimedia systems. Flexible Query Answering Systems, Proceedings 4027, 331–342.

International peer reviewed publications (A2, VABB):

Moelants, D., Cornelis, O., Leman, M., Gansemans, J., Matthé, T., de Caluwe, R., De Tré, G., Matthé, T., & Hallez, A. (2007). Problems and opportunities of applying data- and audio-mining techniques to ethnic music. Journal of Intangible Heritage, 2 pp. 57–69.

Cornelis, O., De Caluwe, R., De Tré, G., Hallez, A., Leman, M., Matthé, T., Moelants, D., et al. (2005). Digitisation of the ethnomusicological sound archive of the Royal Museum for Central Africa (Belgium). IASA Journal, (26), 35–43.

International peer reviewed conference proceedings (C1)

Oramas, S., Cornelis, O. (2012). Past, Present and Future in Ethnomusicology: the computational Challenge. Proceedings of the 13th ISMIR Conference, October 8th-12th, 2012, Porto, Portugal.

Cornelis, O., & Six, J. (2012). Sound to Scale to Sound, a Setup for Microtonal Exploration and Composition. In: Proceedings of the 2012 International Computer Music Conference (ICMC 2012). Ljubljana, Slovenia.

Cornelis, O., & Six, J. (2012). Towards the tangible: Microtonal Scale Exploration in Central-African Music. In: Proceedings of the 2012 Conference for Analytical Approaches for World Music.

Six, J., & Cornelis, O. (2012). A robust Audio Fingerprinter based on Pitch Class Histograms - Applications for Ethnic Music Archives. In Proceedings of the Folk Music Analysis Conference (FMA 2012), Seville, Spain.

Six, J. & Cornelis, O. (2011). Tarsos a Platform to Explore Pitch Scales in Non-Western and Western Music, Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011). Utrecht, The Netherlands.

Cornelis O., Moelants D., Leman M. (2009), Global Access to Ethnic Music: the next big Challenge? Proceedings of the 10th ISMIR, Kobe, Japan. http://www.columbia.edu/~tb2332/fmir/Papers/Moelants-fmir.pdf

Moelants, D., Cornelis, O., & Leman, M. (2009). Exploring African Tone Scales. In Proceedings of the 10th International Symposium on Music Information Retrieval (ISMIR 2009), Kobe, Japan, 2009, pp. 489-494.

Demey M., Cornelis O., Leman M. (2008). The IPEM_EME: a Wireless Music Controller for Real-time Music Interaction, in: Proceedings ICAD 2008, Paris, France.

Cornelis O., Demey M., Leman M. (2008). EME: a wireless Music Controller for Real-Time Music Interaction, Proceedings ARTECH 2008, Porto.

Antonopoulos, I., Pikrakis, A., Theodoridis, S., Cornelis, O., Moelants, D., Leman, M. (2007). Music Retrieval by Rhythmic Similarity applied on Greek and African Traditional Music, in: Proceedings of Eighth International Conference on Music Information Retrieval (ISMIR2007), Vienna, Austria, 23–30 September 2007, pp. 297–300.

Moelants D., Cornelis O., Leman M. (2006). Problems and Opportunities of Audio-mining on Ethnic music, Proceedings of the 7th International Symposium on Music Information Retrieval (ISMIR 2006) pp. 334–336.

Matthé, T., De Tré, G., Hallez, A., De Caluwe, R., Leman, M., Cornelis, O., Moelants, D., et al. (2005). A framework for flexible querying and mining of musical audio archives. 16th International Workshop on Database and Expert Systems Applications, Proceedings (pp. 1041–1045). Presented at the Sixteenth International Workshop on Database and Expert Systems Applications, Los Alamitos, CA, USA: IEEE Computer Society.

International peer reviewed conference abstracts (C3)

Cornelis, O., & Six, J. (2012). Revealing and Listening to Scales from the Past; Tone Scale Analysis of Archived Central-African Music Using Computational Means. In: Proceedings of the 2012 Conference for Interdisciplinary Musicology (CIM 2012).

Six, J. & Cornelis, O. (2011). Tarsos Automated Tone Scale Extraction Applied to African Oral Music Traditions, In : International Workshop of Folk Music Analysis (FMA2011), Athene, Greece.

Lesaffre, M., Cornelis, O., Moelants, D., & Leman, M. (2009). Integration of Music Information Retrieval Techniques into the Practice of Ethnic Music Collections. Presented at the Unlocking Audio 2 - connecting with listeners, London, UK: British Library.

Book chapter (B2)

Cornelis, O. (2011). Theoretisch en artistiek onderzoek naar de symbiose van Westerse en niet-Westerse muzikale idiomen, situering. ARIP: Artistic Research In Progress (pp. 90–99). Gent: School of Arts.

Cornelis, O. (2010). Een digitaal klankarchief, en nu? Over de opportuniteiten van een digitale omgeving aan de hand van het digitaliseringsproject. Achter de muziek aan: muzikaal erfgoed in Vlaanderen en Nederland (pp. 280–284). Acco, Uitgeverij.

Cornelis, O., & Gansemans, J. (2006). De problematiek van het DEKKMMA-project. In M. Leman, O. Cornelis, & A. Ganzevoort (Eds.), Digitale bibliotheken voor muzikale audio: perspectieven en tendensen in digitalisering, archivering en ontsluiting van muziek (pp. 23–35). Presented at the Contactforum Digitale bibliotheken voor muzikale audio, Brussel: Koninklijke Vlaamse Academie van België voor wetenschappen en kunsten.

Book editor (B3)

Leman, M., Cornelis, O., & Ganzevoort, A. (Eds.). (2006). Digitale bibliotheken voor muzikale audio: perspectieven en tendensen in digitalisering, archivering en ontsluiting van muziek. Brussel: Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten.

Table of Contents

Cornelis, O., Lesaffre, M., Moelants, D., & Leman, M. (2010). Access to ethnic music: Advances and perspectives in content based music information retrieval. Signal Processing, 90(4) pp. 1008–1031.

Signal Processing 90 (2010) 1008–1031

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Access to ethnic music: Advances and perspectives in content-based music information retrieval

Olmo Cornelis [a,*], Micheline Lesaffre [b], Dirk Moelants [b], Marc Leman [b]

[a] University College Ghent, Hoogpoort 64, 9000 Gent, Belgium
[b] Ghent University, Blandijnberg 2, 9000 Gent, Belgium

ARTICLE INFO

Article history:
Received 1 December 2008
Received in revised form
23 April 2009
Accepted 11 June 2009
Available online 24 June 2009

Keywords:
Ethnomusicology
Music information retrieval
Access
Music archives

ABSTRACT

Access to digital music collections is nowadays facilitated by content-based methods that allow the retrieval of music on the basis of intrinsic properties of audio, in addition to advanced metadata processing. However, access to ethnic music remains problematic, as this music does not always correspond to the Western concepts that underlie the currently available content-based methods. In this paper, we examine the literature on access to ethnic music, while focusing on the reasons why the existing techniques fail or fall short of expectations and what can be done about it. The paper considers a review of the work on signals and feature extraction, on symbolic and semantic information processing, and on metadata and context tools. An overview is given of several European ethnic music archives and related ongoing research projects. Problems are highlighted and suggestions of the ways in which to improve access to ethnic music collections are given.

1. Introduction

Access to music can be defined as the link that exists between a search idea (“I want music from Rwanda”) and the delivery of a concrete audio file, along with its associated contextual information. Digital computers and wireless networks thereby provide the infrastructure for accessing music from anywhere at any given moment. On top of that, research on music information retrieval (MIR) focuses on tools that extract intrinsic properties from musical audio, related to pitch, rhythm, timbre and so on. These tools enable a large variety of search strategies directly referring to the audio-musical content, in addition to text-based descriptions.

In a recent paper, Casey et al. [1] provide an overview of the advances of audio-based feature extraction and classification methods applied to Western classical and popular music. In the present paper, we address the peculiarities and the difficulties of handling ethnic music. The main problem with ethnic music is that it does not always correspond to the Western concepts that underlie the currently available content-based methods. Therefore, it is of interest to focus on the reasons why the existing techniques fail or fall short of expectations and what can be done about it. Our primary focus will be on ethnic music in European collections, and on European projects that manage these collections[1]. In general, however, this review does not aim to give an exhaustive overview of paper publications, projects and ethnic music collections. Rather, the review aims to illustrate a general problem with digital access to ethnic music by means of relevant examples.

* Corresponding author.
E-mail addresses: olmo.cornelis@hogent.be (O. Cornelis), Micheline.Lesaffre@ugent.be (M. Lesaffre), Dirk.Moelants@ugent.be (D. Moelants), Marc.Leman@ugent.be (M. Leman).

[1] The restriction to the European context is justified by the fact that Europe has a huge heritage of ethnic music collections. Moreover, the problems encountered in European projects are, we believe, exemplary for similar projects outside Europe.

0165-1684/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2009.06.020

The structure of the paper is as follows. Section 2 provides a general rationale of why digital access to ethnic music is problematic. Section 3 provides the background information to trends in ethnomusicology and how these relate to MIR. In Section 4 a quantitative survey of the literature on retrieval of ethnic music is presented. In Section 5 a detailed overview of existing methods for content-based retrieval is given, and Section 6 shows the application of these methods in ongoing scientific projects. Section 7 contains a discussion of some of the major lessons to be learned from it and guidelines for further research are extracted from it. A conclusion briefly recapitulates the main achievements of the current research.

2. Access to ethnic music: a challenge

Access to ethnic music faces different problems. The first problem concerns the particular focus of ethnic music archives. Indeed, ethnic music was (and is) not collected for the purpose of distribution and consumption, but rather for the purpose of preservation and research. Hence, preservation is considered the most important goal, followed by digitization, access, and finally distribution/exploitation. Related to this focus on preservation is the fact that ethnic music is stored in archives that typically contain different collections, often recruited from different field expeditions, using a variety of audio carriers and metadata description methods. Although efforts to digitize these collections are increasing, most of material is still only available on their original carriers. In that sense, the development of tools for access to ethnic music has to take into account the particular focus on preservation and research, in addition to distribution and consumption.

The second problem concerns the tension between the private (industrial) versus governmental (institutional) context. Western popular and classical music has been developed in close interaction with industrial activities, technological innovations and societal developments. In contrast, ethnic music forms part of an oral culture that has only a limited link with commercial and industrial activities. Living ethnic music still does not have a direct access to the digital world, but only gets 'sampled' if a researcher decides to add some information to his or her institutional collection and makes some field recordings. Ethnic music is largely a museum artefact, collected as "silent memory" of oral culture, for the purpose to preserve vanishing traditions, and stored in national institutions. Given that context, institutions from different countries have different policies on presentation and access. They approach ethnic music from different cultural perspectives, and national interests. They use different technological infrastructures, and develop different opinions on how to deal with ethnic music. The MIR community should become aware of the concrete situations as well as policies concerning digitalization of and access to ethnic music archives.

The third problem concerns the access link itself. Search and retrieval of ethnic music requires strategies that deal with a large variability of music, users, search intentions and expectations about the retrieved information, which is fundamentally different from music that is conceived within the established Western standards. Considering the metadata descriptions, the standard approach tends to classify pieces by the names of performer(s) and composer(s). However, in ethnic music, the names of performers and composers are often not available. Instead, what is available is (local) names of instruments, ethnic background of the performers, the time and place of recording or some details about the function of the music. Content-based audio of Western music can be described by concepts such as octave equivalence, division of the octave in 12 equal steps, a harmonic-tonal pitch system, pre-defined metric sets and so on. However, for ethnic music, there is not such a fixed description system that can be applied to all the pieces in a collection. There is no guarantee for octave equivalence, scales can look very different from Western scales and the time organization can be devoid of any regular metric schemata. Hence, finding a suitable content-based description can be problematic, and even the concept of music retrieval (as such) is not without its problems. Indeed, what does it mean to search for music in, and retrieve music from, a culture where the word "music" exists only in connection to its function and the actions involved (such as dancing for rain)? The idea of separating sound from the rest of its physical environment (movement, smell, taste, colour) may well be a peculiar "invention" of the West. The situation can be compared to an approach that would provide access to Western multimedia art through sound only. Consequently, the search is limited, and what is retrieved may be equally restricted.

These problems illustrate that the development of access methods to ethnic music are challenging, especially when compared with Western popular music [1]. There is a strong need for governmental support in digitization and access, as ethnic music is not driven by industrial dynamics. Furthermore, a digital platform is needed that allows access and exchange of information about many different types of ethnic music. Finally, there is a need to reconsider the underlying musical assumptions of audio-based content extraction, which are currently based on Western musical concepts. There is indeed a risk that technologies might restrict access to a kind of music that can be easily processed (e.g. commercial popular music), and thus avoiding the development of tools for ethnic music, which is considered to be too complex, unknown, or does not correspond to Western music concepts that underlie these technologies. In the end, the use of MIR techniques could further marginalize non—Western musical traditions. The combination of digitization and commercial large-scale distribution could push 'vulnerable' music even more into oblivion. Should the MIR-community require that researchers take ethnic music collections into account when developing tools, thus involving a larger variety of contexts, including different cultures?

Digital access to ethnic music provides a huge challenge for music research, and for the MIR community

in particular. Yet, if successful, or even partly successful, ethnic music, owing to its diversity and radical human character, may become an important incentive for innovation in content-based MIR and digital music libraries.

3. Background in ethnomusicology and audio-based taxonomy

3.1. Ethnomusicology

3.1.1. Ethnic music in the strict and broad sense

Within the field of ethnomusicology, music can be divided into three distinct categories, namely (1) *ethnic music* (in strict sense of the word), which comprises the music in cultures without written tradition, (2) *non-Western classical music*, the religious and court music of cultures with a written musical culture, such as China, India or the Middle-East, and (3) *folk music*, which is music from cultures with a written musical culture that does not belong to the classical tradition. In the latter category we find studies on Western folk music, but e.g. also on Japanese folk songs.

In this paper, we will use the term "ethnic music" in the broad sense, including the three categories mentioned above. Distinctions between ethnic music in strict sense, non-Western classical music and folk music will be made if necessary. Work on commercial popular music or even non-Western (e.g. Chinese or Brazilian) popular music or Westernized styles of 'world music' are not taken into account in this paper.

3.1.2. Comparative musicology and ethnomusicology

From a historical perspective, European folk music was the first of those three categories that received some scientific attention, although the early researchers mainly focused on its texts [e.g. 2]. The scientific study of non-Western music started only at the end of the 19th century [3–5], and it was first called 'comparative musicology' (or 'Vergleichende Musikwissenschaft' in German). As the name suggests, the main goal was to compare the different musical styles in order to find musical universals and to establish an evolution in music. After World War II, the term 'ethnomusicology' was introduced by Kunst [6], and this term has been adopted ever since. This term refers to ethnic music in strict sense of the word [7], but there is a general agreement that the study of folk music and non-Western classical music can also be included in this term [8].

3.1.3. The possible link with MIR

In contrast to traditional musicology, which, until recently, relied almost entirely on score representations, ethnomusicology started from field recordings that had to be analyzed and transcribed. Ethnomusicologists had to develop tools to gain insight in unfamiliar musical structures and used techniques like 'phonophotography' [10] to be able to study details of the performances. These efforts can be seen as early examples of audio-based content-extraction technology for music. Interestingly, the central idea of comparative musicology, namely to compare and classify music, is closely linked to tasks performed in MIR-research. Famous examples in this regard are the Sachs-Von Hornbostel tree structure to classify musical instruments [11], the development of a transcription system that allows an easy comparison between different kinds of music [12], or the 'Cantometrics' project aimed at a classification of cultures according to the features of their vocal music [13]. Another link with the current MIR research is the interest in using audio analysis tools [9,14] and audio-based content-extraction tools that allow mass extraction of musical features from large audio databases, including very precise calculations of musical parameters not necessarily related to the Western musical framework.

3.1.4. Computational ethnomusicology

Tzanetakis relates ethnomusicology with computation approaches [9], stating that: (1) there is a need for collaboration between ethnomusicologists and technicians to create an interdisciplinary research field; (2) large audio collections are important to develop automated approaches using machine learning techniques that essentially require large amounts of data for training, which are manually annotated by experts; (3) domain specific techniques are necessary to study the particular characteristics and constraints of different musical styles; (4) content and context interfaces enabling active preservation and enhanced accessibility for non-specialists users should be developed; (5) capturing of body movement during the process of making music by using sensors placed on body and or instrument, provides a new methodology that provides interesting perspectives that could be interesting to apply to ethnic music. The latter somehow acknowledges the fact that audio might not be sufficient to describe and/or retrieve ethnic music.

3.2. Taxonomy for content-based MIR

Linking music with ways to access music is a complex problem because it involves the connection between sound and meaning. In an attempt to decompose this complex problem into more manageable sub-problems, Leman et al. [15] propose an analysis framework of how users may specify musical sound in terms of verbal descriptors. The analysis consists of five levels of description, namely, signal, frame-based, parameter-based, event/gesture-based and concept-based representations. These descriptors have roots in acoustical properties of the musical audio, but they can be connected to higher-level semantic descriptors. Based on this approach [16] further refined this analysis framework into a taxonomy for content-based MIR (Fig. 1).

In this taxonomy, low-level descriptors define content that is closely linked to the acoustical or sensorial properties of the audio signal, typically based on a frame-based analysis of the acoustical wave, such as frequency, duration, spectrum, intensity, and on features associated with sensorial properties, such as fundamental frequency, pitch deviation, periodicity pitch, onset, offset,

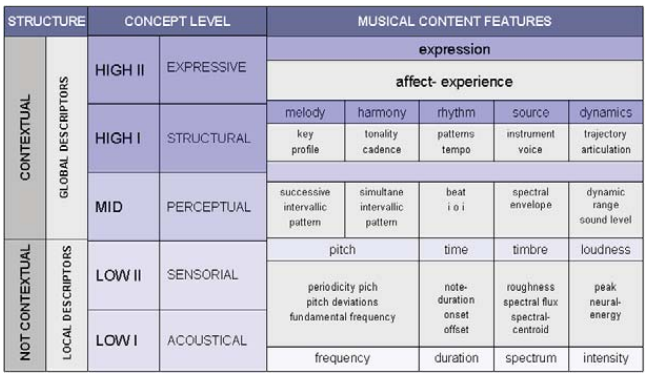


Fig. 1. Taxonomy for content-based MIR.

duration, spectral centroid, spectral flux, roughness, peak neural energy, and others.

Mid-level descriptors involve time–space transformations and context dependencies within specified timescales. Time–space transformations allow for the specification of the musical signal content in spectral terms (e.g. timbre, pitch, chords) and temporal terms (e.g. beat, meter, rhythmic pattern), using time frames of about 3 s [17] to capture representations of what listeners consider to be "the musical present". In that sense, mid-level concepts relate to the so-called "musical parameters" of pitch, time, timbre, and loudness. Related features include simultaneous and successive intervallic patterns, beat, inter-onset-interval, spectral envelope, and dynamic range.

High-level descriptors result from learning and categorization and they typically deal with musical structure and interpretation, these structural descriptions can be further associated with higher-level semantic (e.g. expressive, affective, mood, motor, emotional) descriptors at the highest level [18,19].

3.3. Access categories

The above taxonomy provides a starting point for handling three main access categories, here called, (i) *the signal/feature content*, which comprises the above features of the taxonomy provided that they are extracted from the audio signal, (ii) *the symbolic/semantic content*, which provides information about the meaning of structural and semantic musical features, obtained by manual annotation and associated knowledge-based processing, and (iii) *the metadata/context content*, which provides information about the context of the recording, such as track title, duration, artist genre, and style [1].

It is important to mention that signal/feature content can be seen from two angles, namely, the objective (or physical) angle, and the subjective (or human) angle. The objective angle refers to the measurable feature of sound, i.e. its structural features, whereas the subjective angle refers to what is meaningful in musical sounds. Linking between measurable elements and meaning is a difficult task because it highly depends on subjective interpretation. Lesaffre [20] argues for using a fully integrated music information retrieval system that incorporates a multi-level framework.

Table 1
Number of ISMIR papers related to ethnic music, subdivided by subgenre: (Western) folk music, ethnic music in strict sense of the word and non-Western classical music.

Year	Total papers	Papers dealing with ethnic music	Western folk	Ethnic	Non-Western classical
2000	35	0	0	0	0
2001	43	0	0	0	0
2002	62	3	2	0	3
2003	56	1	0	0	1
2004	108	7	2	3	2
2005	119	6	2	0	4
2006	99	3	0	2	1
2007	131	9	6	3	2
2008	111	9	4	4	3
Total	**686**	**38**	**16**	**12**	**16**

Note that some studies apply to more than one subgenre.

Table 2
Overview on the number of papers spread over three MIR strategies.

Year	Metadata-context	Symbol-semantic	Signal-features
2002	1	1	1
2003	0	0	1
2004	0	4	3
2005	1	2	3
2006	0	0	1
2007	1	5	4
2008	1	4	6
Total	**4**	**16**	**19**

4. Quantitative survey of ethnic music in ISMIR proceedings

The starting point for this survey are the proceedings of the conferences of the International Society for Music Information Retrieval (ISMIR), which are organised every year since 2000 (http://www.ismir.net/proceedings). The main focus of this analysis is on papers that had ethnic music as their main research subject, as well as the papers that had a clearly marked subset of ethnic music in their test set.

4.1. Quantitative survey of ISMIR papers

Table 1 shows the amount of ISMIR papers written over the last nine years. There is a slight increase in the amount of papers dealing with ethnic music, over the last few years, but in total only 5.5% of all the papers deals with ethnic music in some way. Table 2 shows the distribution of all ethnic related papers according to the content level (signal/features, symbol/semantic, metadata/context). There are only few papers that deal with metadata. Instead, there is an almost equal amount of papers about symbol/semantic and signal/feature content.

Table 3
Overview on division of subjects about the MIR topic. The numbers refer to the ISMIR papers listed in Appendix A.

Melody	Harmony	Rhythm	Articulation	Source
3 7 10 24 28 29 32 33 38	5 22	9 11 13 14 16 21 28 29 30 38	34 36 37	
Pitch		Time	Loudness	Timbre
8 12 22 23 31 32 33		4 9 26 28 31 32 37 38		
Frequency		Time stamps	Intensity	Spectrum
3 6 8 10 12 15 20 22 30 31 32 34		4 8 9 13 15 21 23 28 31 32 34 36 37 39		15 30 39

Table 4
Topics outside Table 3 (feature extraction). The numbers refer to the papers listed in the Appendix A.

Concerning metadata	1 17
About the needs of ethnic MIR	2 10 18 19 20 25 35
Presentation of a database	5 18 24 35
Genre classification	3 15 31 39
Recommendation	30 31 33 39
Similarity research	7 11 14 27 29 33 38
Transcription	4 6 8 10 23 24
Performer identificaton	34 36

4.2. Topics of ethnic music papers dealing with musical feature extraction

Table 3 visualises which papers can be catalogued by which musical features. This table is organized according to Fig. 1, the taxonomy of Lesaffre et al. [16]. The horizontal axis represents different "musical parameters". The vertical axis ranges from low-level features (frequency, duration, intensity, spectrum) to high-level features (melody, harmony, rhythm, articulation, source) and an intermediate level (pitch, loudness). The numbers refer to a list of papers that can be found in Appendix A.

The main categories are related to pitch, rhythm, and meter. Only two papers deal with harmony. Features based on intensity, loudness, articulation, spectrum, timbre, and source are less well covered. Although intensity is never a main research topic, it is of course one of the building blocks of onset detection, which is often used in rhythm-related applications.

4.3. Other ethnic music related topics

Table 4 shows papers that focus on other topics than musical feature extraction. Here we find papers in which the construction of a database with ethnic music or the metadata content fields is considered, or papers that call for more specialised MIR research for ethnic music. Some papers dealt with genre classification and similarity research, and there is a recent interest in recommendation systems and performer identification.

4.4. Observations and tendencies

Older papers focused more on annotation and description of one musical feature [21–24], while more recent papers use extracted signal-based content as a step towards higher-level processing, such as recommendation, genre classification, similarity research [25–28]. This trend may be in line with the MIR approach to Western music, in which over the course of several years, the pattern matching paradigm (based on feature extraction and subsequent classification) has been used as the main paradigm for signal-based MIR [29].

It can also be observed that many studies focus on a rather limited audio set, typically based on a selection of a particular musical style, a particular musical instrument or a geographic region. Only a few studies have the ambition to develop tools that could work for a broad range of styles of ethnic music. Seven papers focus on one single musical instrument [23,30,31–35], while most other papers focus on music from a specific geographical entity, sometimes even additionally restricting it to a certain musical style. This reflects a trend in traditional ethnomusicology, which nowadays also tends to focus its research on case studies rather than providing a broader, comparative view. The use of limited datasets may be a sign that signal-based MIR is still in an early stage, involving exploration and trials. Other papers focus more on particular features. For example, many papers deal with Indian music, but instead of focusing on the general problem of database access, the focus is on particular features of the music, such as on labelling tabla percussion [30], pitch contours of Carnatic music [22], pitch analysis of the santur [31], transcription of sitar playing [33], pitch analysis for raga classification [32], or recommendation of Indian classical music [27].

Low-level signal descriptors (e.g. based on short-time magnitude spectrum, constant Q/mel spectrum, pitch frequency annotation, onset-detection, mel/log freq cepstral coefficient, spectral flux, and decibel scale) have been applied to ethnic audio. The global and statistical properties of these descriptors are useful for analysing audio that bears no relationship to Western musical concepts. However, the interpretation of these low-level features in terms of their musical meaning presents a main challenge in this case.

5. Description and representation of ethnic music

In next section, we identify some of the major approaches used to describe and represent ethnic music and problems that go along with this.

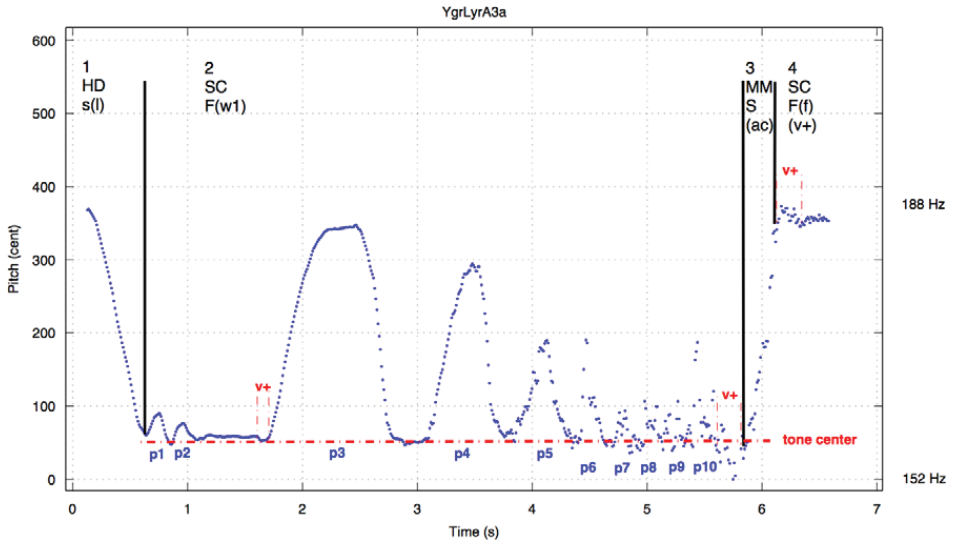


Fig. 2. Example of a guqin tone with rhythmic vibrato and fluctuation patterns. The vertical access shows the pitch modulation in cents, while the horizontal axis shows the time.

5.1. Signal/feature

5.1.1. Pitch processing

The ISMIR publications reflect the fact that pitch analysis (next to issues regarding timing) forms an important research topic. Of the 19 papers that focus on signal features, nine deal with pitch analysis, while two papers included pitch analysis within a compound set of features for genre classification [25,28]. (Also in MIR research applied to Western music, pitch analysis and higher-level classification of chords and keys, has attracted a lot of attention [1]).

However, the further we distance ourselves from the Western music style, the more we are confronted with the fact that the Western concept of pitch, conceived as a discrete category, is too limited or even inappropriate. According to the current MIR approach, a tone is often conceived as an object with a single pitch. The latter can then be represented as a note and further processed from a symbolic point of view. However, music does not always consist of discrete pitch categories. For example, in classical Chinese guqin music [36,37], the music consists of tones that contain sliding patterns (pitch modulations), which form a substantial component of the tone (Fig. 2). The tones are plucked by one hand and modulated by the other hand using a succession of gestures, which results in pitch modulations of the plucked tone (Fig. 3). Accordingly, [37] described the music as a succession of prototypical gestures, applied to a certain tone, rather than as a mere succession of pitches.

Even in European folk music, these pitch gestures can be an important expressive element. In Flamenco singing for example, pitch slides and vibrato play an important role. Gomez and Bonada [39] look for these elements in the music, as an extra element to the pitch detection, but if the results are used for a classification system for flamenco styles [40], they work with simplified structures in which these gestural aspects are left out. This phenomenon is often found in current MIR research on classification of ethnic music. One acknowledges that there are differences, but one uses a reduction to a chromatic scale in order to be able to apply similarity analysis and classification techniques. Chordia [32] for example, starts from recordings of Indian classical music, but as the first step in his raga-classification system, he reduces the pitch content to an equally tempered 12-tone scale.

The second problem concerns the basic tuning system onto which pitches are mapped. In musical cultures like Central-Africa and Indonesia, it appears that scales have an equally-tempered pentatonic structure. Although this would theoretically result in equal intervals around 240 cents, it turns out that the actual tunings varies greatly from one place to another, and the actual scales can even deviate considerably from an equal division of the octave! Another example is the use of a very complex set of scales (maqam) in musical cultures of the Middle-East. These scales contain intervals of different sizes, often not compatible with the chromatic scale, but rather using quartertones between the chromatic steps. These tuning systems are so different from the Western system, that different strategies for pitch analysis have to be developed. In his study of Carnatic (South-Indian classical) music, Krishnaswamy [22,41] introduced a set of 2D melodic units, called melodic atoms (Fig. 4). They are not bound by a regular scale but used as standard patterns that are being matched with the music in order to transcribe, represent or synthesize any melodic phrase.

Bozkurt [42] developed a system to classify Turkish maqams in audio files. He worked with overall frequency histograms to characterize the maqam, separately from a fixed-pitch framework.

While the former two studies focussed on a repertoire with a fixed-pitch framework, the study by Moelants and co-workers on pitch distributions of African music has dealt with a large diversity of irregular tuning systems [43,44]. Using a pitch extraction algorithm [45] to make precise pitch annotations, they avoid creating a priori pitch categories by using a quasi-continuous rather than a discrete (intervallic) representation. Fig. 5 shows some examples of the pitch representations. They allow classi-

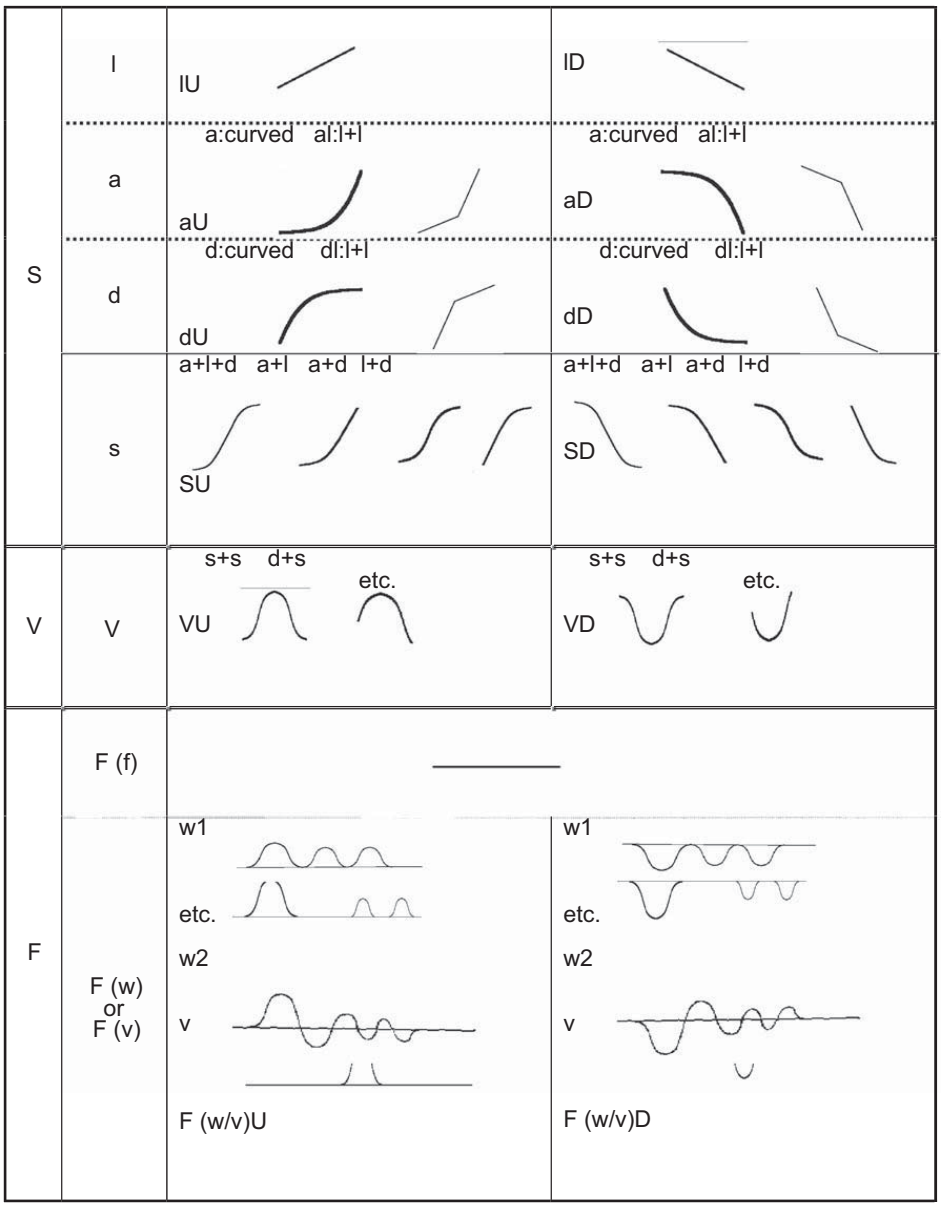


Fig. 3. Tone components (representing portamento and vibrato) from sliding-tones in Chinese guqin music considered as gesture [37].

fication of and comparison with all types of music that use a fixed scale.

5.1.2. Temporal organization

Also the temporal organization of ethnic music can be fundamentally different from the Western standards. A first issue to be aware of is the relationship between meter and rhythm. Instead of working with regular accentual patterns that define a meter, several musical cultures work with patterns that have an irregular rhythmic structure. The application of metric grids (binary and ternary) onto these temporal structures is possible, but it may lead to a simplification of the real underlying rhythmic structures. The detection of the relevant rhythmic patterns often implies knowledge of the specificities of the time organization in the musical repertoire one is studying. For example, Indian music works with a complex multilayered metric system in which the basic units (bols) are connected with syllables. These units each have their own specific sound characteristics, which can be recognized and labelled using the syllables as a first level of analysis [30]. Classical Chinese guqin music does not work with a fixed metrical system, but the gestures that underlie the temporal structure are clearly rhythmical [37].

Another issue that can be problematic in the study of temporal structures in ethnic music is ambiguity. Traditionally, tempo and meter detection systems start from the assumption that every piece of music can be classified in one single category, according to meter or tempo. However, already in Western pop music, the perception of the beat by different listeners can be different [46,47]. In some styles of ethnic music, ambiguity is an essential part of the rhythmical structure, and the interpretation of the meter is connected with associated body movements [48,49].

Several MIR researchers have addressed the above topics. Pikrakis and co-workers [24,26] use self-similarity

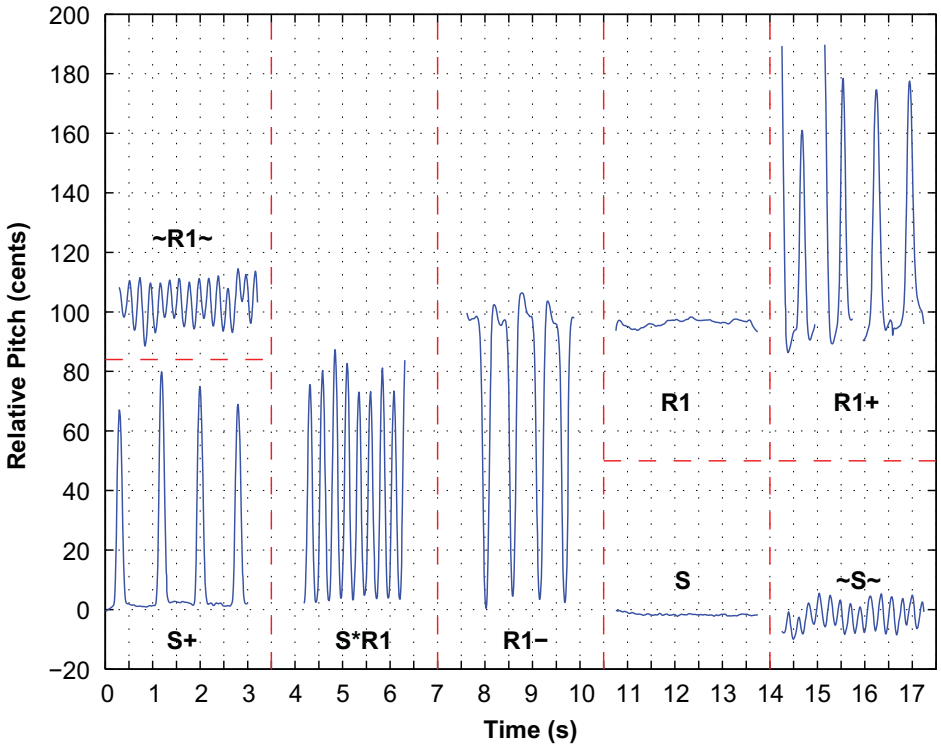


Fig. 4. Typical pitch segments extracted from Carnatic music by Krishnaswamy (2004).

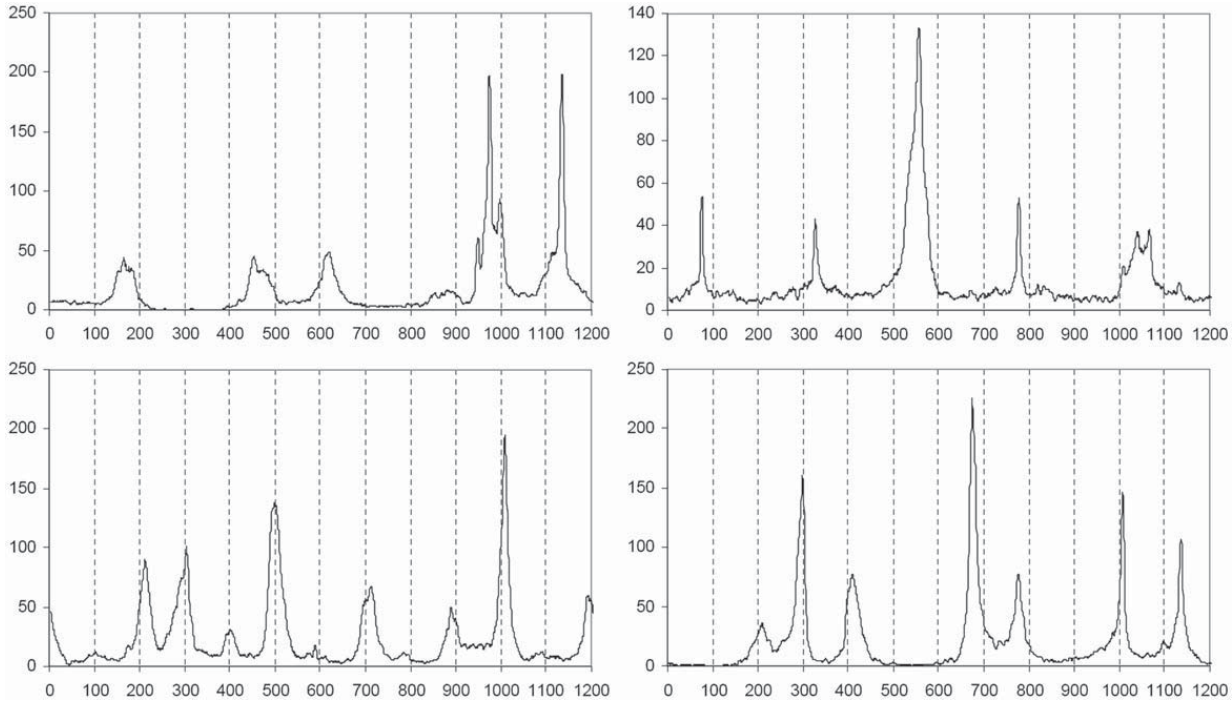


Fig. 5. Four examples of (octave reduced) pitch analysis, from [44]. On the *x*-axis, the pitch with 0 = a, on the *y*-axis the number of occurrences for every 1 cent interval, smoothed by taking the average of seven bins around the middle. The vertical lines show the position of the Western standard pitches, every 100 cents. The upper two graphs represent the pitch distribution in two examples of music from Rwanda: left a song with ikembe (thumb piano) accompaniment and right an umwirongi (flute) piece. The lower two are (left) a Mozart piano sonata (G-major) and a piece of Persian santur music.

matrices as part of their technique for determining and selecting repetitive structures in music. Thus they do not start from a fixed binary-ternary framework, which also allows them to classify irregular rhythmic structures, such as asymmetric meters found in Greek popular music or African rhythmic cells. Wright and Tzanetakis [39] introduce a template-based method for tempo tracking in Afro-Cuban music, which starts from the typical rhythmic pattern found in the claves. The proposed visualization is based on the idea of the rotation-aware

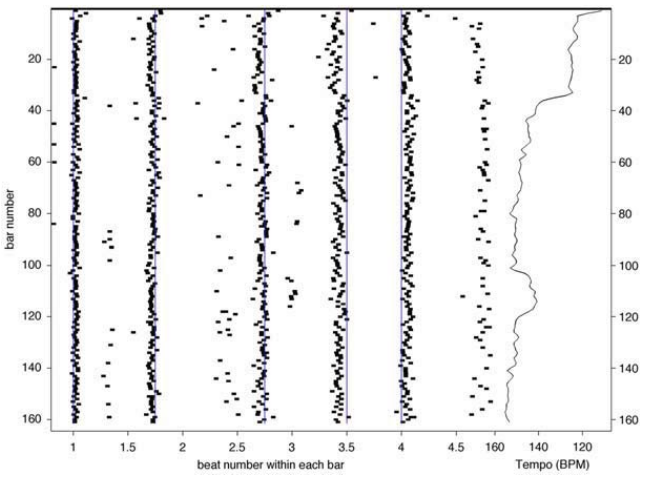


Fig. 6. Bar wrap notation of a rhythmic rumba pattern. The figure depicts the time onsets organised per bar. The straight lines show the theoretical clave locations. The 4th clave note is consistently slightly earlier than the theoretical location, while the 5th clave note is consistently slightly later than the theoretical location.

Bar Wrapping, which is the breaking and stacking of a linear time axis at a fixed metric location. Each bar is then stretched horizontally achieving an equal scaled length of the percussive pattern showing clearly all deviations of the rhythm on micro-level (Fig. 6). This approach is suitable for the analysis of the micro-timing of claves in rumba, but cannot be transferred easily into other styles of ethnic music or onto other musical instruments, as some conditions on which the method relies (such as the consistent timbre of the instrument, the constant and persistent rhythmic pattern and the reoccurring solo introduction of the claves in each song) are unique to a specific style of music.

5.2. Symbolic/semantic

Compared to the previous section, where the sonic particularities of ethnic music form a serious challenge to content-based MIR, the work on symbolic data gives an entirely different picture. Indeed, the availability of huge amounts of symbolically encoded folk songs has presented an opportunity to test for large-scale data-mining capabilities. In some cases, this has lead to the integration of ethnic music collections in the field of advanced MIR research.

5.2.1. Transcriptions of ethnic music

Ethnic music transcriptions should be considered as reductions of original audio performances. Sometimes, these audio performances still exist, but sometimes they no longer exist, or have never existed (e.g. when transcriptions were made by ear, in direct contact with the performers). However, whatever their origin, transcriptions of ethnic music are often problematic. Though Western standard music notation has often been used for comparison purposes only [12], some early ethnomusicologists found Western notation already too restrictive. In the beginning of the 19th century, Villoteau [50] introduced symbols to indicate thirds of a tone in addition to

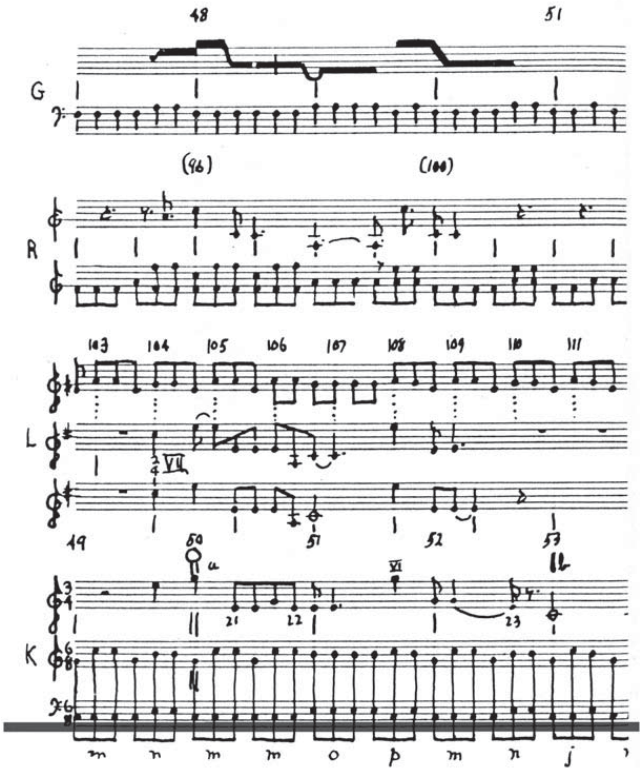


Fig. 7. Transcription of a San (Bushman) song with musical bow accompaniment by four ethnomusicologists (Robert Garfias, Willard Rhodes, George List, and Mieczyslaw Kolinski) at the SEM symposium on transcription and analysis in 1964. The differences in approach illustrate the difficulty of creating a uniform symbolic representation of music that does not use the Western concepts and categorizations.

sharps and flats in his study on Egyptian music. Bela Bartók transcribed music using symbols for small pitch deviations and ornamental variation. Other ethnomusicologists abandoned the use of Western standard notation in favour of a more graphical 'pitch-in-time' representation [e.g. 51]. Therefore, whether the audio collections exist or not, one should always be critical about transcriptions because they face the problem of subjective interpretation. Fig. 7 shows that this can lead to very different representations of the same music, depending on the person who transcribed the music. The problem remains equally critical when working with automated systems, as this forces us to design the analytical strategy for a complete collection [52].

Nevertheless, and apart from the intrinsic problem of transcription of ethnic music, large symbolic databases of (mainly) Western folk music have been created. The size of the collections and the fact that the music is reduced to melodic lines in Western notation make them suitable as testing tools for classification and similarity matching.

5.2.2. Classification

The ISMIR-survey (Section 3) shows a trend in which melody analysis and classification of European folk songs is predominant. The interest in melody classification goes back to the 19th-century nationalist movement and early 20th-century European ethnomusicology. A famous example is the 'Corpus Musicae Popularis Hungarica', based

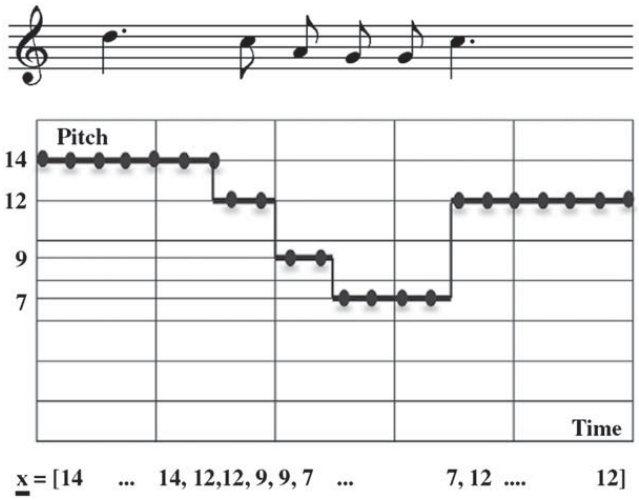


Fig. 8. Melodic contour representation (based on [57]).

on fieldwork by Bela Bartók, Zoltán Kodály and others. From the 1960s onwards, several studies systematically examined this collection [53–55], as well as the idea of digitization and computer-based analysis [56] (see [57] for more details). Meanwhile, this collection contains some 200,000 melody variants. Another example is a work on German folk songs, based on the ESAC code [38,58–60], which is frequently used as a test set, along with the collections of Danish, Finish, or Dutch folksongs [61]. These are but a few examples of large datasets of symbolic ethnic music. Research has hereby mainly focused on melodic representation and classification.

For example, Juhász [62] uses a melodic contour representation for folk music, based on the construction of an n-dimensional vector. Fig. 8 shows how a continuous pitch-time function (shown as a thick line) can be sampled. This can be done vertically by means of integer values that increase one by one according to a semitone division, and horizontally by means of n samples that index the time. The result is stored in a music vector, which allows us to compare melodies independently of their measure, tempo, and syllabic structure. For sampling, it was found that $n = 32$ resulted in an appropriate accuracy for each melody. Because of this, this number was considered the dimension of the "music space".

Using principal component analysis, inherent data structures of folk music can be mapped out as spatial regularities of the point system generated by the melodic mapping technique. For example, Fig. 9 shows the map of the first sections of 2,323 Hungarian folksongs, and possible vertical and horizontal clusters. Each point in the cloud represents a melodic contour. In the figure, some points are linked with melodic contour types and their associated symbolic annotations. The analysis reveals particular rules related to the use of specific intervals (typically tonic, dominant or supertonic at beginning and ending). The approach has been used for the study of variation in the Hungarian oral culture [63], leading to the hypothesis that the basic driving force behind the development in such a culture is selective variation inside the given structures, much more than the composition of radically new, unusual melodies. In that sense, variability contributes to the long-time survival of the oral culture.

Another approach, used to cluster European folk music, is based on self-organisation [57,64,65]. Juhász used a corpus of several thousands of melodies [66,67] to investigate the classification of six different European music cultures, namely Hungarian, Slovakian, French, Sicilian, Bulgarian, and Appalachian English melodies. Using the contour representation (Fig. 9) it was found that the different "music languages" build by the self-organizing maps show a significant overlap, which suggests a "common language" to these cultures, revealing an interaction between European musical traditions and music traditions in the Carpathian Basin. Using self organising maps, it was possible to study songs, on the basis of their interval characteristics, in relation to geographical maps [64], revealing musical traditions and tendencies in Finland (Fig. 10).

5.2.3. *Similarity matching*

Similarity matching is a related relevant research domain in which symbolic collections of both ethnic music and Western music are often used. Similarity forms the core of the pattern matching engines that compare queries (a given symbolic melody) with targets (the symbolic music collection). Similarity may depend on information that involves knowledge of the musical culture. At present, however, there is little difference in the way Western melodies and ethnic melodies are being treated [1,68,69]. In recent work on similarity research, manual annotation has been used to train computational methods [70–72].

Symbol-based MIR research explores a number of symbolic features in view of ethnic music access. Meter and rhythmic information have been shown to be useful in this respect. For example, [73] included melodic information to improve the performance of a meter perception algorithm using an autocorrelation-based meter induction model, [74] focuses on rhythmic similarity as part of melodic similarity measure, and [75] compares rhythmic similarity on the basis of how much insight it provides about the rhythm patterns itself. Distance matrices are determined by the structural inter-relationships that exist within families of rhythms, represented by phylogenetic trees. These trees can be used to show natural relations between patterns, especially when the essence of the rhythm (or its origin) has been simplified. An alternative notation is the chronotonic representation [76]. [77] also presents a novel method for representation, called Metrical Circle Map, allowing us to explore the cyclic aspects of musical time. In the field of pitch representation, Uitdenbogerd [78] encounters problems with representing music which uses different tuning systems as well as with finding matches to queries of this microtonal music. Therefore the Micro-Tonal Representation for Information Retrieval (MTRI) was designed as a music representation method useful for microtonal representation. It consists of the essential elements required for melody, harmony, and rhythm representation, which is sufficient to capture the recognisable elements of a piece of music.

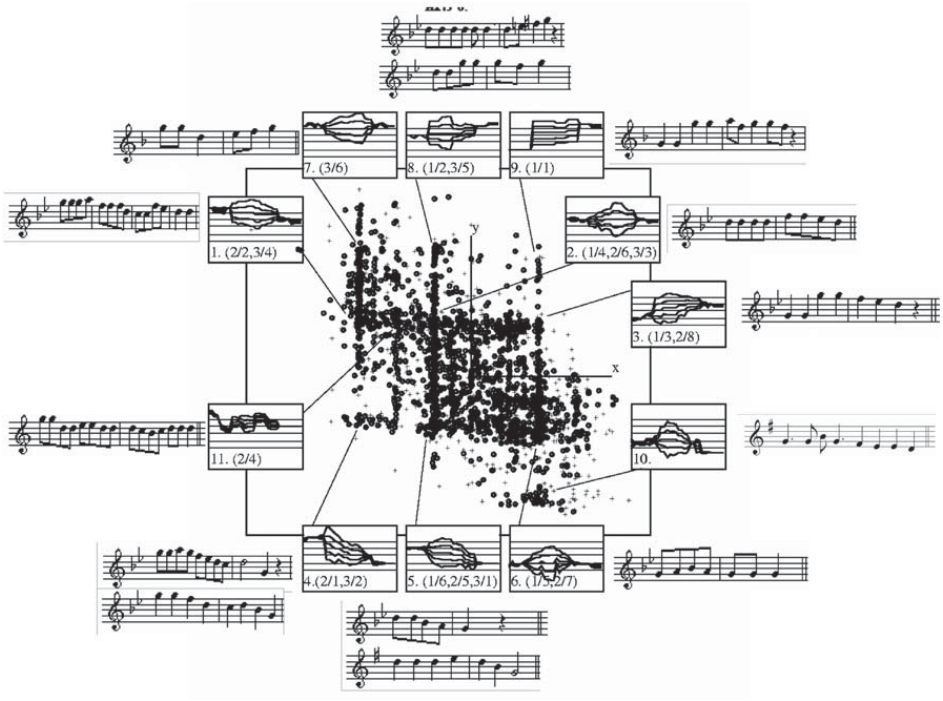


Fig. 9. Clustering using PCA.

An exhaustive overview of symbol-based methods is far beyond the scope of the present paper. It suffices here to say that the above examples illustrate that symbolic ethnic music collections have come into the picture mainly because some big symbolic (European folk) music collections provide datasets onto which statistical classification techniques can be straightforwardly applied.

5.3. Metadata and context

The work on metadata for ethnic music is a reflection of the broad historical and cultural diversity in which these collections have been developed. Relevant problem areas are: the connection of audio signals with metadata, the different platforms for metadata and the use of metadata descriptors beyond the usual Western categories.

5.3.1. Connecting signals with metadata

A collection of ethnic music typically consists of audio files and metadata, which provide information about the context of the recording, such as the geographic and ethnic origin of the music, the recording circumstances, the type of instruments, the function of the music and so on. The link between the audio signal and its proper metacontent is essential for its retrieval and presents a more urgent problem than in Western music, where composers or performers can be more easily identified.

The BWF (Broadcast Wave Format) allows a connection between audio (in WAV format) and the associated minimum metadata that is considered necessary for all broadcast applications [79]. However, it cannot contain entire sets of metadata, and has no direct search options. Audio files and their metadata descriptions have a tricky relationship, especially in the field of ethnographic recordings. If the connection between the audio and the metadata is lost, it is indeed very hard to re-retrieve the exact metadata because of the limited availability of the objects (no web tagging possible on the signal domain), the resemblance of many songs, and the variety of similar performances. Historical reference can be made to the famous sound archive (founded by Stumpf) at the Museum für Völkerkunde Berlin, where after World War II, it turned out that wax cylinders were stored in East Berlin, and the metadata in West Berlin, a situation that persisted until the German re-unification [80].

5.3.2. Platforms for digital metadata

There are several platforms available for storing digital metadata. However, no single metadata standard for music covering all requirements, is available at the moment [81] (Table 5).

The Dublin Core Metadata Initiative is usually seen as the most "scientific" approach to metadata, although the specific needs of ethnic music metadata description are not always fully taken into account. Metadata descriptions for ethnic music can vary a lot because of the changing recoding contexts, leading to a large variety in both quantity and quality of metadata descriptions. Metadata fields are described by a data dictionary that defines all musical metadata terms and categories, such as description of the item, administrational issues, structure, legal rights, and technical information of the remediation [82,83]. Examples are given in Tables 6 and 7.

5.3.3. Ethnic metadata in the DEKKMMA sound archive

In what follows, we propose a more detailed view on metadata organisation for ethnic music, based on our experience with the DEKKMMA project. The focus of this project is on the digitization of and access to audio and

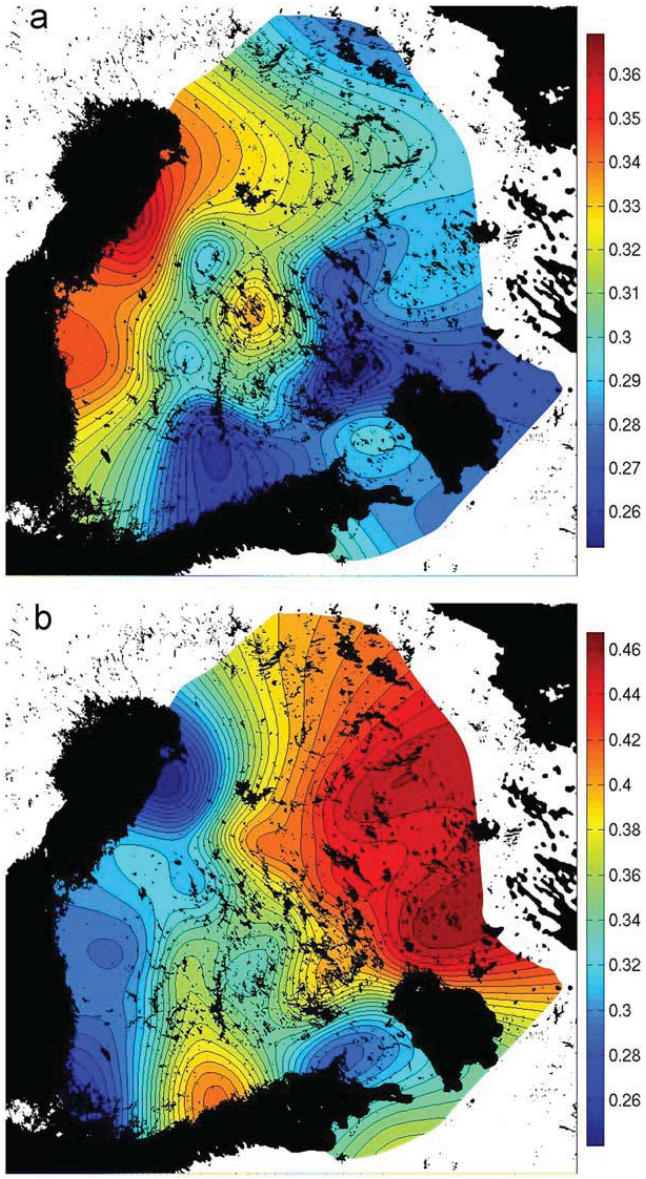


Fig. 10. Musical Features of Finnish folk music related to geographical information. Fig. 10a shows proportion of tunes starting with the tonic. Fig. 10b shows proportion of tunes in minor key.

metadata of the music collection of the Royal Museum for Central Africa in Tervuren, Belgium. Fig. 11 shows the Enhanced Entity Relationship diagram (EER) that summarizes the organisation of the audio database and its metadata. The EER diagram was created by first collecting all metadata fields from the paper index cards and afterwards establishing all dependencies between those metadata fields. The resulting EER diagram visualises three main groups of metadata: (i) data on identification of the tape (such as number of the tape, original carrier, reproduction right, collector, date of recording, duration, (ii) ethno-geographical information (country, province, region, village, people, language), and (iii) data on musical content (style, participants, functions, instrumentation).

However, as a rule, the metadata organisation of ethnic music depends on the institution's history. Ethnomusicological field work results in a variety of description systems, influenced by the personal background and interest of the researchers. For example, the oldest field recordings of ethnic music (outside Europe), which date from the late 19th century, were often made by the military or by priests. They did not follow a strict methodology, but rather collected all kinds of 'exotica'. Often the focus was on the material object (the musical instrument) rather than the immaterial object (the musical audio and its relevant context). In later ethnomusicological research, the structure and content of the annotations also highly depended on the object of study and the scope of the research.

Another aspect worth mentioning here is the relativity of the geographical information over time. Nomination of regional, national and even international idiom is often related to the current administration, and it may easily change after periods of conflict or reorganisation of administrative practice. The complexity of the terminology used in ethnomusicology is also pertinent in the labelling of musical instruments. We can illustrate this with the example of the thumb piano or lamellophone in the database of the DEKKMMA project. This type of instrument appears in 1290 audio recording, spread over 26 countries. The large amount of vernacular names (see Table 8) reflects a great variety of languages and the different typologies of the musical instrument. A lot of these vernacular names are closely related. Many of these vernacular names show clearly morphological relationships, resulting from age-long cultural inheritance. However, some background knowledge is necessary to relate all these terms to the same type of instrument.

To sum up, the history, the large variety of metadata, and the changing local conditions still have an impact on the way modern archives have to deal with ethnic music. It is important to realise that the large variety in metadata fields of ethnic music, both at formal and content level, is a reality, and that, therefore, the organisation of an audio database and its metadata descriptions should be developed in close collaboration with engineering and (ethno)-musicology (Table 9).

5.3.4. *Flexible querying of ethnic metadata*

The DEKKMMA project also explored the use of fuzzy logics, as a way to deal with certain peculiarities of the metadata [85]. Traditional query systems work with a three-valued propositional logic ($T=$ True, $F=$ False, $\wedge=$ Not applicable), and an output that is limited to one of these three levels. In contrast, Extended Possibilistic Truth Values (EPTVs) allow a more flexible search in databases, making it possible that database entries which, for example, satisfy most (but not all) of the constraints will also be present in the query result [84]. This makes it easier for users to find what they are actually looking for in a database.

For ethnic music collections, flexible querying may be a very promising means to deal with missing, vague or incomplete metadata, as well as with differences in spelling. Typical problems relate to geographical locations, time inaccuracies and vernacular names of instruments. In annotations of ethnic music, often the date or place are not exactly determined. Instead we can only find that a recording was made, for example, around 1970 and/or

Table 5
Overview of the potential of several metadata schema's by Corthaut [81].

	ID3	freeDB	MusicBrainz	Dublin core	Music vocabulary	Music ontology	MPEG-7	MusicXML
Musical info	+	+/−	+/−	−	+/−	+	+	+
Classifiers	+/−	+/−	−	+/−	+/−	+/−	+	+/−
Performance	+	+/−	+/−	+	+	+	+	+/−
Versioning	+	−	−	+	+/−	+	+	−
Descriptor	+	+	+	+	+	+	+	+
Rights and ownership	+	−	−	+	−	+	+	+
Playback rendition	+	−	−	−	−	+	+	+
Lyrics	+	−	−	−	+	+	+	+
Group. and refs.	+	−	−	+	+/−	+	+	+
Identifiers	+	+	+	+	−	+	+	−
Record-info	+	+	+	−	−	+	+	−
Instr. and arrang.	+/−	−	−	−	+	+	+	+
Sound and carrier	−	−	−	+	−	+	+	+/−
Event	−	−	−	−	+	+	+	−
Time-modelling	−	−	−	−	−	+	+	+
Notation	−	−	−	−	+/−	+/−	−	+
Attention-metadata	+	−	−	−	−	+	+	−
Publishing	+	−	−	−	−	+	+	−
Composition	+	−	−	+/−	+	+	+	+
Production	−	−	−	−	−	+	+	−
Meta-metadata	−	+	−	−	−	−	+	−

Table 6
Data Dictionary developed by McGill University, 2006.

Field name	trackDuration
Definition	Given the duration of a track
Multiplicity	0…1
Data type	Record in the representations of ISO 8601:hh:mm:ss
Label	Track duration
Example or notes	00:02:59
Data constraints	Track level
Version tracking	Issued on 2004-07-13

Table 7
Data dictionary developed by Harvard University Library, 2004.

Number	1
Name	Duration
Definition	The length of sound in time-code character format (TCF)
Required	M
Repeatable	N
Values	Any valid time-code character format value
Mapping	
Examples	00:10.25.14
Notes	TCF may require the use of indicator codes '<' and '&', which cannot be entered directly in XML format

nearby Kinshasa. By introducing a degree of constraint satisfaction, the given query results will also contain elements that did closely match the query. For instance, the closer to the date 1970, or the closer to Kinshasa, the higher the constraint satisfaction degree will be. This allows to search beyond the strict limitations of the search terms.

Furthermore, one can provide the users with the extra possibility of attaching a greater importance to some of the selection criteria for calculating the output of the query [85]. For example, the user may find it more important that the music was played with a particular musical instrument than that the recording took place "around 1979". Values between 0 and 1, attached to the selection criteria, can be used to acknowledge the respective importance of the different criteria, where a value of 1 denotes "absolute importance" and value 0 denotes "not important at all". Finally, the results can then be ranked according to their calculated satisfaction grades, hereby giving the user an indication of which results match (globally) the best with the imposed selection criteria [86].

5.3.5. Recent trends

During the ISMIR 2008 conference, methods for tagging music by web crawling on available user-contributed content metadata were presented by several contributors. Obviously, this is a scenario in which the quality of the result depends on the quantity of the available data. Relying on the gigantic efforts of users tagging the music, automatic web-based services may result in a flexible and large music recommendation system for music listeners. However, such an approach is currently not available for ethnic music. Metadata for ethnic music are very particular, often specialized and difficult to access. Tagging methods may push ethnic music even more into oblivion [44]. As stated by Chordia [27], metadata on Indian classical music is often missing or inaccurate, and user tagging of Indian classical music tracks is uncommon.

However, massive tagging may still provide a new way of accessing ethnic music, if, based on how non-experts speak about this kind of music, for example, by focusing on semantic appreciation and mood-related descriptions. It is not clear whether this idea is realistic or not because mood related descriptions may be very different among listeners, depending on their own cultural background. Moreover, it is not at all easy to label ethnic music with subjective parameters. On the other hand, providing metadata that do not refer to aspects of musical structure, should allow non experts to find their way to alternative music more easily. In this context, Fingerhut [87] has

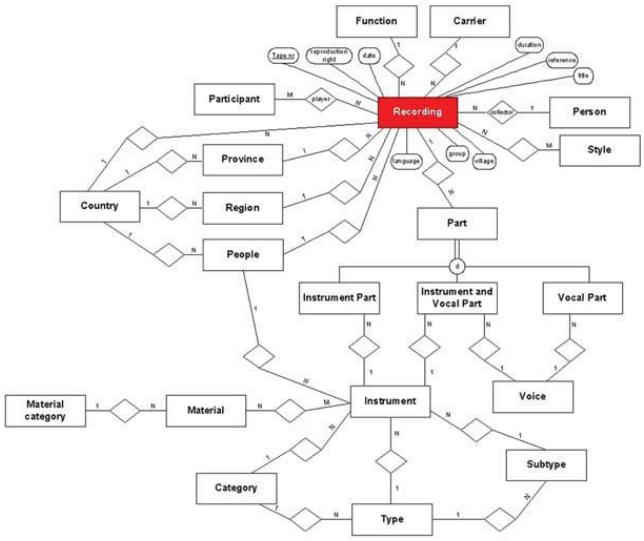


Fig. 11. EER diagram of the metadata model used in the DEKKMMA project.

Table 8
Variation in vernacular names of the lamellophone.

Tshisaasj	Ngombi
Tshisaji	Kombi
Tshisanji	Kembe
Chisanzi	Ekembe
Sanzi	Ikembe
Sanza	Dikembe
Sansa	Likembe
Esanzo	Kalimbe
Issanji	Malimba
Kisanzi	Marimba
Kassandji	Irimba
Kassayo	Ilimba
	Limba
	Kadimba

Table 9
Exemplary overview of possible EPTV values assigned to seven database items if 1970 was queried.

Date	EPTV
1970	(T,1)
Around 1970	(T,1),(F,0.3)
Unknown date	(T,1),(F,1)
1973	(T,0.8),(F,0.4)
1910	(F,1)
N/A	(^,1)
Inconsistent	(T,0.5),(F,1),(^,1)

The user gets not only the exact matches to his query, but (s)he can also search in (closely) related songs. Assigned penalties for deviating results are manually predetermined.

pointed to the problem of navigating within a single audio document.

6. Collections and projects

The complexity of ethnic music archives (see Section 2) makes it extremely difficult to design a unified approach to access music collections. The reasons for this complexity are explained by its history [79,88,89] and by the fact that ethnomusicological archivists and researchers have to cope with hybrid collections, which are partly analogue and partly digital. Such archives require the management of traditional documents together with modules that cover all the aspects of digital documents. Moreover, what do users want and what can be provided for different user categories? These are difficult questions that need careful consideration even before starting the development of any access tools.

6.1. Ethnomusicological collections in Europe

Based on web browsing, checking IASA membership (International Association of Sound and Audiovisual Archives), and consulting publications on ethnic music, it is possible to get a non-exhaustive overview of the state of the art in digitization and access to ethnic music collections.

However, such a survey is not without its problems. The first problem that occurs is that ethnic music collections are often part of larger institutions (e.g. British library, KMMA Tervuren) that deal with a lot more than just music, such as video recordings, films, documents, and artefacts (such as music instruments, masks, cloths, and so on). This may be one of the reasons why a universal / commonly agreed upon exhaustive list of ethnic music collections is not available, nor is there an exhaustive list available that summarises their state of the art in digitization and in content-based access. The second problem is that almost all major ethnic music archives in Europe use digital catalogues but few of them make the catalogues available online. Smaller collections are often not able to offer modern access services. Apart from these collections, it is likely that quite a number of smaller, private, or other collections are missing from our list. Media collections maintained by radio and television broadcasters, for example, have not been taken into account because ethnic music forms mostly a very small part of their collection. The third problem is that access to digital collections is often limited to educational purpose, due to the fact that services depend on problems with digital rights management [90]. It is often the case that solutions for IPR (Intellectual Property Rights) are inadequately developed even though they are crucial in meeting the users' demands.

6.2. Survey of ethnic music collections

To obtain more information about the state of the art in digital access to ethnic music collections, a questionnaire was sent to 44 institutions that we found through our web survey as described in 6.1 (see list in appendix). The questionnaire aimed at gathering information about the description of the respective archives in terms of size, content, carriers, period, and region. Furthermore questions were asked about the metadata, digitization, use of standards, interoperability with users, rights management, ethical issues, and access policies. The study aimed at getting a small glimpse of the real world of digital

access to ethnic music collections in Europe of which there must be many hundreds. Out of the 44 institutes that we addressed barely 7 (16%) reacted to our request. Although the questionnaire was rather short, more than half of the respondents did not provide answers to all questions. We received answers from small and medium-sized collections in archives and in one multimedia library. Four of the seven collections entirely consist of ethnic (i.e. mainly folk) music, within the other three collections ethnic music forms rather a small part. One collection is wholly digitized, two others have been partly digitized, for another collection plans for digitization are made up and the remaining did not start digitizing yet. Used standards for metadata are MARC, UNIMARC and OCLC PICA. The standards used for the audio are wav-files (archive) and mp3-files (publishing). From the 4 collections that provide interoperability with users, 2 are restricted to researchers, 1 is free but requires registration and acceptance of terms and 1 has a digital rights management system implemented at the kernel of the information system.

The poor response to our questionnaire is indicative for the field. Furthermore, small heritage collections often suffer from financial, theoretical and technical problems. Newly developed techniques used to handle ethnic music are not always within the reach of non-experienced staff members. The challenge of working with heterogeneous material has been particularly taxing. Each genre requires a slightly different approach in terms of how it is documented and which elements need to be added to clarify the presentation of the audio.

6.3. Current situation of research projects in Europe

In the past decade, an increasing number of European projects (both national and EC-sponsored) has been set up in the area of digital libraries and archives (e.g. DELOS, DISMARC, MICHAEL, kopal, nestor, PrestoSpace, TAPE, EASAIER) (for abbreviations used in this section, see Appendices II and III). The projects cover mixed collections of audiovisual documents as well as other media of which ethnomusicological material forms often just a small part. Only a few projects address questions regarding the preservation, accessibility, connectedness or exploitation of ethnomusicological cultural heritage (e.g. EthnoArch, DEKKMMA, Pofadeam).

6.3.1. Projects related to mixed collections

The DISMARC project (Discover Music Archives across Europe, http://www.dismarc.org/) project started in 2006 and was set up to support the uncovering of large amounts of under-exposed European cultural, scientific, and scholarly musical audio collections by increased exposure/use of content, multi-lingual simultaneous searching, and the creation of a free, on-line content catalogue. All interested archives can participate in the DISMARC project by allowing their content to be searchable via the DISMARC search engine. DISMARC produced a metadata template, through which any archive or collector can present content in the DISMARC portal. By using the metadata template, a mapping procedure can be followed which presents any kind of data set in an accurate manner in the DISMARC portal, regardless of its original format. The DISMARC portal has been online since September 2008. At present, DISMARC combines content from 57 various archives, including ethnic material (e.g. from the EMEM, ISPAN, MAM and DFA collections). There are currently around 1.3 million records, including 2000 audio tracks, on the server. The audio functionality is to be developed further as DISMARC becomes the audio pillar of EUROPEANA, the European digital library. Several controlled vocabularies have been developed, which support the search and ingest process of the technical system, namely for Genres, Formats, Geography, Eras, IBSS (International Bibliography of Social Sciences), Ethnic/Non-ethnic groups, Languages, Partner Countries. These vocabularies are offered as SOAP/WSDL-based WebServices and allow export to EUROPEANA in SKOS-Format. DISMARC search engine users can search for specific items such as composers, collections, archives and labels. The audio player provides a random selection of material from the DISMARC collection. All audio is immediately playable, and details about the artist and the owner of the particular piece are provided. A demo tool shows DISMARC's search possibilities. At present only a small selection of DISMARC audio is connected to the DISMARC World Map (http://www.dismarc.org/info/audio_map.html). This map enables the user to browse through audio by zooming into a country. For example, in its current state, the DISMARC map for Europe found 126 songs from 2 archives representing 1 genre (Fig. 12). A zoom into France for instance retrieves only one song, a Vietnamese traditional song "Ve Mien Trung", performed by Tran Quang Hai, stored at the EMEM archive. Audio discovered via the DISMARC search engine can be added to a basket for licensing if required. A modular licensing contract between the would-be user and the owner of the particular item will be created on the spot and the parties will be put in contact with each other.

The ASR project (Archival Sound Recordings, http://sounds.bl.uk/) is part of a multi phase (ASR1 and ASR2) digitization programme (2003–2009) of the Joint Information Systems Committee (JISC) that is funded by the UK Higher Education and Further Education funding bodies. The focus is on the innovative use of ICT to support education and research. The ASR website offers over 1500 recordings. For copyright reasons, the other recordings can only be played in licensed UK higher and further education institutions or in the library's reading rooms. Among the collections in the ASR project, are Divid Rycrofts Africa recordings (unaccompanied choral singing, songs composed for indigenous musical instruments, and urban music), 200 rare wax cylinder recordings from the World and Traditional Music collections of the British Library Sound Archive and Klaus Wachmanns recordings of indigenous music in Uganda. Though available for listening, the recordings in some particular sections, such as ethnographic wax cylinders are only freely available as streaming audio files and not for download. The first ASR project (ASR1), incorporated various standards relating to digitization of audio and images, the development of the web interface, and the delivery of metadata. All metadata

Fig. 12. DISMARC world map.

are delivered in the XML-based Metadata Encoding and Transmission Standard (METS), containing legacy information about the archival original, documenting the process of digitization and audio segmentation, together with the provision of standard descriptive data encoded in Dublin Core. The descriptive data taken as standard for the ASR project was provided by the specially developed British Library Application Profile for Sound (BLAP-S), an audio-oriented extension of the existing British Library Application Profile. BLAP-S information is embedded in the METS record of each item on the site. The British Library was committed to the adoption of METS as a standard for managing some of its digital archiving processes.

EASAIER (Enabling access to sound archives integration enrichment retrieval, http://www.elec.qmul.ac.uk/easaier/) and MEMORIES (Design for an audio semantic indexation system allowing information retrieval for the access to archive content, http://www.memories-project.eu/) are two EC-sponsored research projects that aim at creating an access system for sound archives, incorporating tools for indexing, and the manipulation of resources [91]. Although the projects do not focus on ethnic music they can be applied to aspects of storing and manipulating ethnic music audio.

6.3.2. *Projects related to ethnomusicological collections*

There are a few projects that focus exclusively on ethnic music archives. The ethnoArc project (Linked European Ethnomusicological Archives, http://www.ethnoarc.org/) is an EC-project that aims to improve access to collections from ethnomusicological archives in Europe (i.e. Bucharest, Budapest, Berlin, and Geneva) by developing a number of tools for database creation, filling, handling, and searching for ethnomusicological archives. The ethnoArc software allows ethnomusicological archives to define their data structure in XML, from which a database is built. Metadata can then be entered and managed. Researchers can define queries (spanning multiple archives) on the metadata. The ethnoArc multi archive search tool (ethnoMARS) allows researchers to build and perform queries on one or more ethnomusicological archives, without the need of having additional information about the archive, such as written documentation or personal contact. The primary purpose of the project is to enable researchers to search available metadata at the archive sites, based on criteria defined by the researcher, and store the results locally for post-processing and reference. At present only metadata from the archives participating in the ethnoArc project can be queried. These archives are AIMP—Archives Internationales de Musique Populaire [Geneva, Switzerland], EMEM—Ethnological Museum Berlin—Department for Ethnomusicology [Berlin, Germany], IEF—Constantin Brailoiu Institute for Ethnography and Folklore [Bucharest, Romania], and ZTI—Institute of Musicology of the Hungarian Academy of Sciences [Budapest, Hungary].

The Dutch national project WITCHCRAFT (What Is Topical in Cultural Heritage: Content-based Retrieval Among Folksong Tunes) (http://www.cs.uu.nl/research/projects/witchcraft/) aims at developing a fully functional content-based retrieval system for folksong melodies stored as audio and notation, using the best practices of MIR research. The WITCHCRAFT project uses the Yet Another Heterophone: Musical Utrecht University Global Lookup Engine (YahMuugle) search engine. The current developmental version of YahMuugle provides content-based retrieval of melodies from the Dutch folk song collection of the Meertens Institute (about 6000 digitized melodies). There are various ways to formulate a query, including using a clickable piano keyboard, humming or whistling in a microphone, playing on a midi device, and using a database item as query. In the list with the results

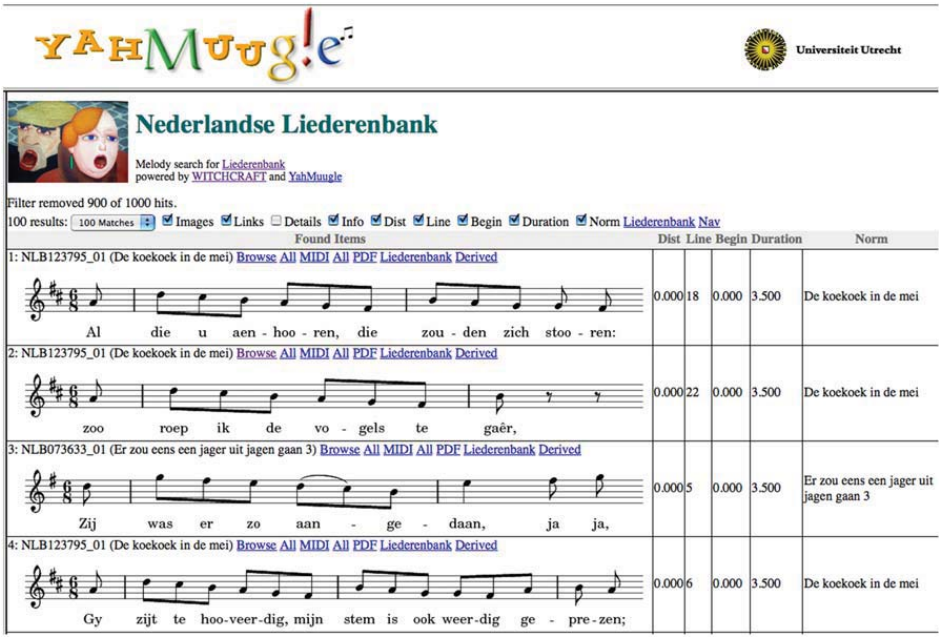


Fig. 13. YahMuugle result list.

the phrase that closely matched the retrieved song is shown, along with some contextual information (Fig. 13). For matching similarities among melodies, a version of the Earth Movers Distance is used.

The WebFolk.BG (WebFolk Bulgaria Global Inventory Project) aims to increase access to integration, application, and use of distributed sources of musical data, and to develop a related technology to support these multimedia resources through Internet/Web. This work won the first award (1997) for a regional Bulgarian project coordinated by Global Inventory Project at the European Commission, G7 and Japan. However, only little information on the WebFolk project is available in the English language.

The Authentic Bulgarian Musical Folklore (ABMF) multimedia database http://musicart.imbm.bas.bg/EN/Default.htm is online available. It integrates folk music data (songs, instrumental music, dances and rituals), information, and related analytical tools. It is based on information found in an old and highly valuable archive of Bulgarian folk music collected for more than 70 years and preserved in the Institute for Art Studies-BAS. This multimedia database was created in 1994. Today there are more than 16,000 records associated with integrated audio, photos, score facsimiles, and video. The database user can choose between several search options, namely search by list, search by map and complex search.

The Belgian national project DEKKMMA (Digitization of the Ethnomusicological Sound Archive of the Royal Museum for Central Africa, http://music.africamuseum.be) digitized the entire sound archive and metadata (37,389 records) of the Royal Museum for Central Africa and developed a database and website for public access, applying flexible querying and content-based MIR techniques.

The structure of the existing index cards provided the basis for the database design, in which user evaluation was involved (see Section 5 on metadata). A distinction has been made between the access functionalities for internal users (those allowed to insert new data and/or update existing data and consult the data) and the access functionalities for external users (all others, only allowed to consult the data). For dissemination, a website has been developed enabling all metadata and parts of the audio to be available. To establish the context of the sound recordings more clearly, additional information about countries, people, and musical instruments has been provided on the website. Currently, the website allows textual search by free text fields, searching by thesaurus lists, and geographical browsing using a map of Africa. To allow cooperation between different archives, metadata is formatted in a standardized way so it can be handled by open archive initiative (OAI), an exchange format that organises the internal representation of the information. DEKKMMA also developed audio-based MIR techniques for automated analysis of large amounts of audio. All audio files have been provided with a pitch and tempo analysis, both in graphical and numerical way.

6.3.3. *Users and user needs*

In the DEKKMMA project a distinction was made between access functionalities for internal users and the access functionalities for external users. Clearly, knowledge about users is an important for the development of proper access portals. Since the migration of online public access catalogues systems (OPACs) to web-based interfaces with video and music, library staff and public can access the catalogue from outside the library (e.g. British Library Sound Archive, Archives Sonores du CREM). In this context, ethnomusicological archives have to meet an increasing demand for consultation from users who expect online access to digital and analogue material as well.

Users today expect immediate digital access to ethnic music, similar to the digital access of Western music via

iTunes or other providers. Given the context of ethnic music, this may imply the fast delivery of MP3 files together with accompanying multimedia documentation about the music's context (e.g. text, photographs, and videos). Therefore, it seems that access to multiple contents of collections through free and well-maintained websites is a possible pathway for the future. It is likely that the provision of enriched information in association with music is a model for future access to music, both ethnic and Western. The latter would rely on audio-based content in combination with an elaborate set of semantic and metadata information.

However, further research is needed to address different potential audiences of different age groups or nationalities, with different interests, and at varied levels of understanding of the field. For example, the EthnoArc project made a distinction between following user types: scholars from various disciplines (ethnomusicology, musicology, cultural anthropology, historiography), students, musicians, teachers, sound engineers, and amateurs. Another classification [43] distinguishes three main groups. A first group is the general public, which has an interest in (some kind of) ethnic music. These users typically want to retrieve music using a rather vague and general labelling, such as 'African drumming', 'trance music' or 'Inuit songs'. A second group consists of users who come from within the culture. They may have a good knowledge of certain repertoires and functions of the music, and therefore, they tend to ask very specific questions about music played by a specific performer, music from one particular village, lyrics, genres, or instruments (in local terminology). A third group are the researchers, who use the archive to enhance their study. This group would typically ask questions about the geographical spread of certain instrument types, the relative importance of certain rhythm, pitch, or musical structures in different regions.

This brief overview shows that user-oriented research is an underdeveloped research area of MIR. To improve access to ethnic music collections (and perhaps to music collections in general) further effort will be needed to develop this domain and to link it with the engineering approaches to audio-content extraction and classification.

7. Discussion and conclusions

Previous studies have called for more culturally independent tools for content-based MIR [9,90,92,93]. In 2002, Futrelle and Downie [92] pointed already to a significant lack of content-based MIR-tools that cover ethnic music. They argue that a radical rethinking of the MIR research practice is needed and concluded with the statement that it is imperative that studies of existing music systems include non-Western resources and their use in non-Western contexts. In 2003, Downie [93] phrased a similar call to the MIR-community to take care of developing techniques for all kinds of music, including non-Western music. Recently Tzanetakis [9] writes that: "Historically, the majority of work in MIR has focused on either popular music with applications such as music recommendation and personalized radio systems or on Western 'classical' music with applications such as score following and query-by-humming, thus dividing clearly a segregation between classical and popular music, just because of their own needs of specific MIR applications". The present review shows that content-based MIR tools are often biased by Western music concepts and that new, more universal, tools for ethnic music are urgently needed. A more close collaboration between engineers and musicologists may be a first step towards achieving that goal.

The present review further shows that the most advanced access tools for ethnic music often rely on a musical reduction to Western score notation (pitch-time events). This reduction to familiar Western musical parameters is evident in the studies that deal with European folk music. We believe that reduction can be very useful for providing high-level access to collections. However, a simplified symbolic representation, which reduces the musical content to a pre-established set of categories and discards any gestural aspects of the music, may lead to a loss of much of the rich and lively texture of this music. A possible solution is to keep signal/feature content and symbolic/semantic content as closely connected as possible. Further research is needed to better integrate culturally neutral audio representations and features [92] with symbolic and meaningful content descriptors.

Integration is also needed between signal/feature content, symbolic/semantic content, and metadata/contextual content. Better integration of these levels may lead to more advanced music recommendation systems that link audio with a rich set of multimedia data. The diversity of ethnic music can be an incentive for renewed content delivery systems in the music industry. In any case, given the diversity of collections, there is an urgent need for platforms that allow connections between the different metadata formats and their linkage with audio and multimedia documents.

This review shows that there are many ways in which access to ethnic music can be improved. For example, existing systems for meter detection (currently limited to binary and ternary meters) can be extended to quintuple or septuple meters. In a similar way, a system for key recognition can be adapted to recognize modes used in European folk music or Indian ragas.

However, a system for meter detection cannot say anything useful about non-metric music for example, and the system for key recognition based on a 12-tone scale cannot say anything about microtonal music or music that does not assume the octave, like Indonesian Gamelan music [94]. Unless the search space can be constrained by a priori information provided by the users, the performance of the MIR system may be less robust if it needs to differentiate between a large number of categories. It is likely that interfaces will therefore increase in complexity, as the user is offered a larger choice. More work is needed in finding optimal solutions to these problems.

One way to approach complex ethnic music collections seems to be the development of culturally and style independent query systems. Examples are existing systems that start from uncategorized user input, such

as query-by-humming or query-by-example. An important advantage of these systems is the avoidance of interpretation and representation of the specific musical characteristics since the user obtains similar recordings as output without really having to define, read, or understand the query results. Query-by-example may in a long term, be the most promising technique to approach complex and large databases, as it allows every possible sound input. However, much work has to be done to enable its application to various characteristics. When dealing with completely different styles of music, the key question becomes what constitutes a 'similarity' [95].

We believe that audio-based bottom-up driven solutions to similarity should be developed together with music-based top-down driven solutions. Audio-based bottom-up driven solutions have the advantage of closely resembling audio. Music-based top-down driven solutions have the advantage of closely resembling culturally determined categories, meaning and experience. The challenge of MIR-research is to combine audio and culture, physics and meaning, to the benefit of all those that create and consume music of all cultures and of all styles.

Appendix A. List of annotated ISMIR papers

See Table A1.

Appendix B. Research projects

See Table B1.

Appendix C. Collections

See Table C1.

Table A1

NR	Author(s)	Title	Year	Level	Collection
1	Blandford A. and Stelmaszewska H.	Usability of Musical Digital Libraries: a Multimodal Analysis	2002	M	Folk Music Collection
2	Geekie G.	Carnatic Ragas As Music Information Retrieval Entities	2002	–	Indian Carnatic Music
3	Tzanetakis G. et al.	Pitch Histograms in Audio and Symbolic Music Information Retrieval	2002	L/S	Irish folk songs
4	Gillet O.K. and Richard G.	Automatic Labelling of Tabla Signals	2003	L	Indian music (Tabla)
5	Cruz F.W. et al.	A Brazilian Popular Music Digital Library	2004	S	Brazilian popular music
6	Krishnaswamy A.	Melodic Atoms for Transcribing Carnatic Music	2004	L	Indian Carnatic music
7	Mullensiefen D. and Frieler K.	Optimizing Measures of Melodic Similarity Of Folk Songs	2004	S	Folksong (Luxembourg)
8	Nesbit A. et al.	Towards Automatic Transcription of Australian Aboriginal Music	2004	L	Australian Aboriginal music
9	Pikrakis A. et al.	Music Meter and Tempo Tracking from Raw Polyphonic Audio	2004	L	Greek Folk music
10	Suyoto I.S.H. and Uitdenbogerd A.L.	Exploring Microtonal Matching	2004	S	Sundanese Songs
11	Toussaint G.T.	A Comparison of Rhythmic Similarity Measures	2004	S	Traditional West-African and Afro-American
12	Heydarian P. and Reiss J.D.	The Persian Music and The Santur Instrument	2005	L	Indian Music (santur)
13	Jensen K. et al.	Rhythm-Based Segmentation of Popular Chinese Music	2005	L	Chinese popular music
14	Kapur A. et al.	New Music Interfaces for Rhythm-Based Retrieval	2005	S	Ethnic hyperinstrument (sitar, drum)
15	Norowi N.M. et al.	Factors Affecting Automatic Genre Classification: an Investigation Incorporating Non-western Musical Forms	2005	L	Traditional Malay music
16	Toivianen P. and Eerola T.	Classification of musical Metre with Autocorrelation and Discriminant Functions	2005	S	Essen Folk COLLECTION, Finnish Folk Tunes
17	Walser R.Y.	Herding Folksongs	2005	M	Madison Carpenter collection (folk)
18	Castro B.M. et al.	Bdb-mus a Project for the Preservation of Brazilian Musical Heritage	2006	–	Several collections of Brazilian music
19	Doraisamy S. et al.	Towards a Mir System for Malaysian Music	2006	–	Traditional Malay music
20	Moelants D. et al.	Problems and Opportunities of Applying Data- and Audio-Mining Techniques	2006	L	Traditional African music from RMCA
21	Antonopoulos I. et al.	Music Retrieval by Rhythmic Similarity Applied on Greek and African Traditional Music	2007	L	Traditional African and Greek Dance Music
22	Chordia P. and Rae A.	Raag Recognition using Pitch-Class and Pitch-Class Dyad	2007	L/S	Indian Raga's
23	Kapur A. et al.	Pedagogical transcription for multimodal sitar	2007	L	Indian sitar music
24	Lanzelotte R. et al.	A digital Collection Of Brazilian Lundus	2007	L/M	19th century Brazilian popular music (lundu)
25	Vankranenburg P. et al.	Towards Integration of Mir and Folk Song Research	2007	–	Dutch Folk Songs

Table A1 (*continued*)

NR	Author(s)	Title	Year	Level	Collection
26	Frieler K.	Visualizing Music on the Metrical Circle	2007	S	Irish folk from Essen collection
27	Garbers J. et al.	Using pitch Stability among a Group of Aligned Query	2007	S	Dutch Folk Songs
28	Pikrakis A. and Theodoridis S.	An application of empirical mode decomposition	2007	L	Greek traditional music
29	Volk A. et al.	Applying Rhythmic Similarity based on Inner Metric	2007	S	Dutch Folk Songs
30	Chordia P. et al.	Extending Content-based Recommendation: the case of Indian Classical Music	2008	L	North Indian classical music
31	Doraisamy S. et al.	A Study on Feature Selection and Classification Techniques for Automatic Genre Classification of Traditional Malay Music	2008	L	Traditional Malay Music
32	Duggan B. et al.	Machine Annotation of sets of Traditional Irish Folk Tunes	2008	L/S	Traditional Irish Folk Tunes
33	Garbers J. and Wiering F.	Towards Structural Alignment of Folk Songs	2008	S	Dutch Folk Songs
34	Ramirez R. et al.	Performer Identification in Celtic Violin Recordings	2008	L	Irish popular music
35	Silla C.N. et al.	The Latin Music Database	2008	M	Latin music
36	Trajano de Lima E.	On Rhythmic Pattern Extraction in Bossa Nova Music	2008	S	Brazilian Popular Music (Bossa Nova)
37	Wright M. et al.	Analyzing afro-Cuban Rhythm using Rotation-Aware Clave Template Matching with Dynamic Programming	2008	L	Afro-Cuban music (Rumba)
38	Volk A. et al.	A manual Annotation Method for Melodic Similarity and the Study of Melody Feature Sets	2008	S	Dutch Folk Songs
39	Yoshii K. and Goto M.	Music Thumbnailer: Visualizing Musical Pieces in Thumbnail Images based on Acoustic Features	2008	S	Latin, African and Japanese music

List shows all papers annotated in Table 3. (Nr—author—title—year—level—collection).
L = low level, S = symbolic, M = metadata.

Table B1

Acronym	Name	Target collections	Focus	Coordinator	URL
ASR	Archival Sound Recordings	M	Education and research	British Library, London	http://sounds.bl.uk/
CEDARS	Curl exemplars in digital archives	M	Preservation	University of Leeds	http://www.leeds.ac.uk/cedars/
DEKKMMA	Digitization of the Ethnomusicological Sound Archive of the Royal Museum for Central Africa	E	Digitization of African music	IPEM, Gent University	http://music.africamuseum.be/
DELOS	Network of Excellence on Digital Libraries	M	Technology/systems	association	http://www.delos.info/
DISMARC	Discovering Music Archives Across Europe)	M	Portal for integration and coordination	consortium	http://www.dismarc.org/
EASAIER	Enabling access to sound archives integration enrichment retrieval (EASAIER)	M	Access and preservation	Queen Mary University, London	http://www.elec.qmul.ac.uk/easaier/
EthnoArc	Linked European Ethnomusicological Archives	E	Access and preservation	Wissenschaftskolleg zu Berlin	http://www.ethnoarc.org/
Kopal	Co-operative development of a long-term digital information archive	M	Technology/archiving system	Niedersaechsische Staats-und Universitaetsbibliothek Goettingen	http://kopal.langzeitarchivierung.de/
Nestor	Network of Expertise in Long-Term Storage of Digital Resources (Nestor)	M	Preservation	Deutsche Nationalbibliothek, Frankfurt am Main	http://www.langzeitarchivierung.de/
MICHAEL and MICHAEL PLUS	Multilingual Inventory of Cultural Heritage in Europe	M	Inventorization	EU countries	http://www.michael-culture.org/en/home
MINERVA eC	Digitizing Content Together	M	Digitization	Ministero per i Beni e le Attività Culturali, Rome	http://www.minervaeurope.org/home.htm
MultiMatch	Multilingual/Multimedia access to cultural heritage (MultiMatch)	M	access	Istituto di Scienza e Tecnologie dell' Informazione, Pisa	http://www.multimatch.eu/

Table B1 (*continued*)

Acronym	Name	Target collections	Focus	Coordinator	URL
MUSIC NETWORK	The interactive MUSIC NETWORK	M	Technology/ systems	Universita Degli Study Di Firenze	http://cordis.europa.eu/data/
POFADEAM	Preservation and On-line Fruition of the Audio Documents from the European Archives of etnic Music	E	Preservation	University of Udine	http://www.ipem.ugent.be/2005POFADEAM/
PrestoSpace	Preservation towards storage and access. Standardized Practices for Audiovisual Contents in Europe.	M	Preservation and access	Institut National de l'audiovisuel, Paris	http://www.prestospace.org/
	Pròiseact Thiriodh	E	Online access	University of Edinburough	http://www.tiriodh.ed.ac.uk/indexe.html
SDM	Spectacles du monde (SDM)	E	Web portal (available in 2009)	Cité de la musique, Paris	www.spectaclesdumonde.fr
TAPE	Training for Audiovisual Preservation in Europe	M	Preservation	ECPA	http://www.tape-online.net/
	Vienna's international music spectrum	E	Audiovisual documentation	Phonogrammarchiv, Vienna	http://www.pha.oeaw.ac.at/
WebFolk.BG	WebFolk Bulgaria	E	Preservation and access	Bulgaria, Sofia	http://musicart.imbm.bas.bg/project_bulfolk.html
WITCHCRAFT	What Is Topical in Cultural Heritage: Content-based Retrieval Among Folksong Tunes	E	Technology/ MIR systems	University of Utrecht	http://www.cs.uu.nl/research/projects/witchcraft/

M = mixed, E = ethnic.

Table C1

Acronym	Collection	Location
AIMP	Archives Internationales de Musique Populaire	Musée d'ethnographie de Geneve, Switserland
AMA	Archiv für die Musik Afrikas	Institutes für Ethnologie und Afrikastudien der Johannes Gutenberg-Universität Mainz
	Archive of Greek Music	Institute of Research on Music and Acoustics (IEMA), Athens
CREM	Archives sonores du CREM	Centre de recherche en ethnomusicology, Paris
	Archivio Roberto Leydi	DECS, Divisione della cultura, Centro di dialettologia e di etnografia, Ticino
EMEM	Berlin Phonogramm-Archiv	Ethnological Museum, Berlin
	Bibliotheca de Catalunya Fonoteca	Bibliotheca de Catalunya, Barcelona
	British Library Sound Archive	British Library, London
	Collectie Jaap Kunst	UVA-Universiteitsmuseum Amsterdam/Instituut Multiculturele muziek Studies (IMS), Amsterdam
DFA	Dansk Folkemindesamling	Copenhagen
DVA	Deutsches Volksliedarchiv	Institut für internationale Popularliedforschung, Freiburg
EP BSAM	The Ethnomusicological Phonoarchive of the Belarusian State Academy of Music	Belarusian State Academy of Music, Minsk
	Essen Folksong Collection	Helmut Schaffrath Laboratory of Computer Aided Research in Musicology, in Warsaw
ERESBIL	Eresbil Musikaren euskal artxiboa	ERESBIL-Basque Archives on Music, Errenteria
KMMA/ RMCA	Sound Archive of the Ethnomusicology of the Royal Museum for Central Africa	Royal Museum for Central Africa, Tervuren, Belgium
EFA	Estonian Folklore Archives	Tartu, Estonia
GTF	Gesellschaft für Historische Tontrager	Phonomuseum, Wien
	Digital Archive of Finnish Folk Tunes (Suomen Kansan Sävelmiä)	Finnish Literary Society, University of Jyväskylä
	A-K collection	Folklife Archives Sound Archives, Department Musicology, Tampere
	Greek Folk Music Archive	Music Library of Greece Lilian Voudouri, Athens
	Audio Collection of the Hungarian Heritage House	Hungarian Heritage House, Budapest
	Irish Traditional Music Archive	Irish Traditional Music Archive, Dublin
IEF	Constantin Brailoiu Institute for Ethnography and Folklore	Academia Romana, Boekarest
ISPAN	ISPAN sound collection	Institute of Arts of the Polish Academy of Sciences, Warsaw
SDM	Médiathèque de la Cité de la musique	Cité de la musique, Paris
	Multimedia Database for Authentic Bulgarian Musical Folklore	Institute of Art Studies, Bulgaria
MAM	Wolfgang Laade Music of Man Archive	Hochschule für Musik und Theater Hannover (HMTH)/University of Music and Drama, Hannover
MEN	Collection du MEN	Musée d'ethnographie, Neuchâtel
	Collections de la Médiathèque du Musée du Quai Branly	Musée du Quai Branly, Paris
	Fons Sonor Museu de la musica de Barcelona	Museu de la musica, Barcelona
	Nederlandse Liederenbank	Meertens Instituut, Amsterdam
	Norsk Musikksamling	Nasjonalbiblioteket, Oslo

Table C1 (*continued*)

Acronym	Collection	Location
	Norwegian Folk- and Popular Song Archives	Norsk Visearchiv, Oslo
	Norwegian Folk Music Collection (Norsk Folkemusikksamling)	Department of Musicology, Oslo University
	Norwegian Institute of recorded Sound	Norsk Lydinstitutt, Stavenger
	Phonogrammarchiv	Austrian Academy of Sciences, Vienna
	Schweizer Nationalphonothek	Swiss National Sound Archives, Lugano
SOAS	SOAS Library Music Collection	School of Oriental and African Studies, London
SRC SASA	Sound Archive of the Institute of Ethnomusicology	Slovene Academy of Sciences and Arts (SAZU), Ljubljana
	Sound Archive of the School of Scottish Studies	University of Edinburgh
SVA	Svenskt Svisarkiv	The Centre for Swedish Folk Music and Jazz Research, Stockholm
ZRC	Archive of the Ethnomusicology Institute of the Slovenian Academy of Sciences and Arts	Slovenska akademija znanosti in umetnost, Ljubljana
ZTI	Archives of the Institute of Musicology of the Hungarian Academy of Sciences	Institute of Musicology of the Hungarian Academy of Sciences, Budapest

References

[1] M.A. Casey, et al., Content-based MIR: current directions and future challenges, Proceedings of IEEE 96 (4) (2008) 668–695.
[2] J.G. Herder, Volkslieder, Weimar, 1778.
[3] A.J. Ellis, On the musical scales of various nations, Journal of the Royal Society of Arts 33 (1885).
[4] C. Stumpf, Lieder der Bellakula-Indianer, Vierteljahrschrift für Musikwissenschaft (1886).
[5] R. Wallaschek, Primitive Music, Longmans, London, 1893.
[6] J. Kunst, Musicologica, Royal Tropical Institute, Amsterdam, 1950.
[7] A. Danielou, Traité de musicologie comparée, Hermann, Paris, 1959.
[8] B. Nettl, Comparative study and comparative musicology: comments on disciplinary history, in: A. Schneider (Ed.), Systematic and Comparative Musicology: Concepts, Methods, Findings, UITGEVERIJ, Frankfurt, 2008.
[9] G. Tzanetakis, et al., Computational ethnomusicology, Journal of Interdisciplinary Music Studies 1 (2) (2007) 1–24.
[10] C. Seashore, H. Seashore, The place of phonophotography in the study of primitive music, Science (1934).
[11] E.M. von Hornbostel, C. Sachs, Systematik der Musikinstrumente: Ein Versuch, Zeitschrift für Ethnologie (1914).
[12] E.M. von Hornbostel, O. Abraham, Vorschläge für die Transkription exotischer Melodien, Sammelbände der Internationalen Musik-Gesellschaft (1909).
[13] A. Lomax, Cantometrics, University of California, Berkeley, 1976.
[14] A. Schneider, Comparative and systematic musicology in relation to ethnomusicology: a historical and methodological survey, Ethnomusicology (2006) 236–258.
[15] M. Leman, et al., Musical audio mining, in: J. Meij (Ed.), Dealing with the Data Flood: Mining Data, Text and Multimedia, STT Netherlands Study Centre for Technology Trends, Rotterdam, 2002, (ISBN 90–804496–6-0), pp. 440–456.
[16] M. Lesaffre, et al., User-dependent taxonomy of musical features as a conceptual framework for musical audio-mining technology, in: Proceedings of the Stockholm Music Acoustics Conference, Stockholm, Sweden, 6–9 August 2003, pp. 635–638.
[17] M. Leman, et al., Correlation of gestural musical audio cues and perceived expressive qualities, in: A. Camurri, G. Volpe (Eds.), Gesture-based Communication in Human–Computer Interaction, Springer, Berlin Heidelberg, 2004, pp. 40–54.
[18] M. Leman, et al., Prediction of musical affect attribution using a combination of structural cues extracted from musical audio, Journal of New Music Research 34 (1) (2005) 39–67.
[19] M. Lesaffre, et al., How potential users of music search and retrieval systems describe the semantic quality of music, Journal of the American Society for Information Science and Technology 59 (5) (2008) 1–13.
[20] M. Lesaffre, Music Information Retrieval: Conceptual Framework, Annotation and User Behavior, 2005, (unpublished Ph.D. Thesis).
[21] G. Tzanetakis, et al., Pitch histograms in audio and symbolic music information retrieval, in: Proceedings of Third International Conference on Music Information Retrieval, Paris, France, 13–17 October 2002, pp. 31–38.
[22] A. Krishnaswamy, Melodic atoms for transcribing Carnatic music, in: Proceedings of Fifth International Conference on Music Information Retrieval, Barcelona, Spain, 10–15 October 2004, pp. 345–348.
[23] A. Nesbit, et al., Towards automatic transcription of Australian aboriginal music, in: Proceedings of Fifth International Conference on Music Information Retrieval, Barcelona, Spain, 10–15 October 2004, pp. 326–330.
[24] A. Pikrakis, et al., Music meter and tempo tracking from raw polyphonic audio, in: Proceedings of Fifth International Conference on Music Information Retrieval, Barcelona, Spain, 10–15 October 2004, pp. 192–197.
[25] N.M. Norowi, et al., Factors affecting automatic genre classification: an investigation incorporating non-western musical forms, in: Proceedings of Sixth International Conference on Music Information Retrieval, London, UK, 11–15 September 2005, pp. 13–20.
[26] I. Antonopoulos, et al., Music retrieval by rhythmic similarity applied on greek and African traditional music, in: Proceedings of Eighth International Conference on Music Information Retrieval, Vienna, Austria, 23–30 September 2007, pp. 297–300.
[27] P. Chordia, et al., Extending content-based recommendation: the case of Indian classical music, in: Proceedings of Ninth International Conference on Music Information Retrieval, Philadelphia, USA, 14-18 September 2008, pp. 571–576.
[28] S. Doraisamy, et al., A study on feature selection and classification techniques for automatic genre classification of traditional Malay music, in: Proceedings of Ninth International Conference on Music Information Retrieval, Philadelphia, USA, 14–18 September 2008, pp. 331–336.
[29] R.O. Duda, et al., Pattern Classification, Wiley, New York, Chichester, England, 2001.
[30] O.K. Gillet, G. Richard, Automatic labelling of Tabla Signals, in: Proceedings of Fourth International Conference on Music Information Retrieval, Baltimore, Maryland, USA, 26–30 October 2003.
[31] P. Heydarian, J.D. Reiss, The Persian music and the Santur instrument, in: Proceedings of Sixth International Conference on Music Information Retrieval, London, UK, 11–15 September 2005, pp. 524–527.
[32] P. Chordia, A. Rae, Raag recognition using pitch-class and pitch-class dyad distributions, in: Proceedings of Eighth International Conference on Music Information Retrieval, Vienna, Austria, 23–30 September 2007, pp. 431–436.
[33] A. Kapur, et al., Pedagogical transcription for multimodal sitar performance, in: Proceedings of Eighth International Conference on Music Information Retrieval, Vienna, Austria, 23–30 September 2007, pp. 351–352.
[34] R. Ramirez, et al., Performer identification in Celtic violin recordings, in: Proceedings of Ninth International Conference on Music Information Retrieval, Philadelphia, USA, 14–18 September 2008, pp. 483–488.
[35] M. Wright, et al., Analyzing Afro-Cuban rhythm using rotation-aware clave template matching with dynamic programming, in: Proceedings of Ninth International Conference on Music Information Retrieval, Philadelphia, USA, 14–18 September 2008, pp. 647–652.
[36] H. Penttinen, et al., Aspects on physical modeling of a Chinese string instrument—the guqin, in: Ninth International Congress on Acoustics, Madrid, 2–7 September 2007.

[37] H. Li, M. Leman, A gesture-based typology of sliding-tones in guqin music, Journal of New Music Research (2007) 61–82.
[38] H. Schaffrath, The essen associative code: a code for folksong analysis, in: Beyond MIDI: the Handbook of Musical Codes Book Content, MIT Press, Cambridge, MA, USA, 1997, pp. 343–361.
[39] E. Gomez, J. Bonada, Automatic melodic transcription of flamenco singing, in: Proceedings of the Fourth Conference on Interdisciplinary Musicology, Thessaloniki, 3–6 July 2008, pp. 66–67.
[40] J.J. Cabrera, et al., Comparative melodic analysis of a Cappella Flamenco Cantes, in: Proceedings of the Fourth Conference on Interdisciplinary Musicology, Thessaloniki, 3–6 July 2008, pp. 38–39.
[41] A. Krishnaswamy, Multi-dimensional musical atoms in South-Indian classical music, in: Proceedings of ICMPC, 2004.
[42] B. Bozkurt, An automatic pitch analysis method for turkish maqam music, Journal of New Music Research 37 (1) (2008) 1–13.
[43] D. Moelants, et al., Problems and opportunities of applying data- and audio-mining techniques to ethnic music, in: Proceedings of Seventh International Conference on Music Information Retrieval, Victoria, BC Canada, 8–12 October 2006, pp. 334–336.
[44] D. Moelants, et al., Problems and opportunities of content-based analysis and description of ethnic music, International Journal of Intangible Heritage (2007) 58–67.
[45] T. De Mulder, et al., Recent improvements of an auditory model based front-end for the transcription of vocal queries, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Montreal, 2004, pp. 257–260.
[46] M. Mc Kinney, D. Moelants, Ambiguity in tempo perception: what draws listeners to different metrical levels?, Music Perception 24 (2) (2006) 155–165.
[47] M. Mc Kinney, et al., Evaluation of audio beat tracking and music tempo extraction algorithms, Journal of New Music Research 36 (1) (2007) 1–16.
[48] J. Phillips-Silver, L.J. Trainor, Vestibular influence on auditory metrical interpretation, Brain and Cognition 67 (2008) 94–102.
[49] L. Naveda, M. Leman, Representation of Samba dance gestures, using a multi-modal analysis approach, in: Proceedings of Fifth International Conference on Enactive Interfaces, Pisa, European Enactive Network of Excellence ENACTIVE, 2008, pp. 68–74.
[50] G. Villoteau, De l'état actuel de l'art musical en Egypte, Paris : Imprimerie de C.L.F Pancroucke, 1826.
[51] B.I. Gilman, Hopi Songs, Journal of American Ethnology and Archaeology (1908).
[52] A. Klapuri, M. Davy, Signal Processing Methods for Music Transcription, Springer, New York, USA, 2006.
[53] P. J'ard'any, Experiences and results in systematizing Hungarian folksongs, Studia Musicologica XXII (1974) 17–20.
[54] J. Szendrei, Auf dem wege zu einer neuen stilordnung der ungarischen volksmusik, Studia Musicologica XX (1978) 361–379.
[55] L. Bobszay, Der begriff typus in der ungarischen volksmusikforschung, Studia Musicologica XX (1978) 227–243.
[56] K. Csébfalvy, et al., Systematization of tunes by computers, Studia Musicologica VII (1965) 253–257.
[57] Z. Juhasz, A systematic comparison of different European folk music traditions using self-organizing maps, Journal of New Music Research 35 (2) (2006) 95–112.
[58] U. Franzke, Das EsAC-Project, in: Typological Classification of Tunes, Advanced Systems for Arranging Folklore Stocks, Lithuanian Academy of Music, Department of Ethnomusikology, Vilnius, 1996.
[59] U. Franzke, Untersuchungen zur Intervallstatistik des deutschen Volksliedes mit Hilfe des EsAC-Systems, in: Typological Classification of Tunes, Advanced Systems for arranging Folklore Stocks, Lithuanian Academy of Music, Department of Ethnomusikology, Vilnius, 1996.
[60] U. Franzke, H. Schaffrath, PAT-Ein Patternsuchprogramm für Essener und andere Software, Handbuch, Universität Essen, 1991.
[61] P. Vankranenburg, et al., Towards Integration of Mir and Folk Song Research, in: Proceedings of Eighth International Conference on Music Information Retrieval, Vienna, Austria, 23–30 September 2007, pp. 505–508.
[62] Z. Juhasz, Contour analysis of Hungarian folk music in a multi-dimensional metric-space, Journal of New Music Research 29 (1) (2000) 71–83.
[63] Z. Juhasz, A model of variation in the music of a Hungarian ethnic group, Journal of New Music Research 29 (2) (2000) 159–172.
[64] P. Toiviainen, T. Eerola, Visualization in comparative music research, in: A. Rizzi and M. Vichi (Eds.), Proceedings in Computational Statistics. Physica-Verlag, Heidelberg, 2006, pp. 209–221.
[65] Z. Juhasz, Analysis of melody roots in Hungarian folk music using self-organizing maps with adaptively weighted dynamic time warping, Applied Artificial Intelligence 21 (1) (2007) 35–55.
[66] K. Csébfalvy, et al., Systematization of tunes by computers, Studia Musicologica VII (1965) 253–257.
[67] L. Dobszay, J. Szendrei, Catalogue of Hungarian Folksong Types, Institute of Musicology, Budapest, 1992.
[68] W.B. Hewlet, E. Selfridge-Field, Music query. Methods, models and user studies, Computing in Musicology 13 (2004) 183.
[69] P. van Kranenburg, et al., Towards integration of music information retrieval and folk song research department of information and computing sciences, Utrecht University Technical Report UU-CS-2007-016, www.cs.uu.nl, ISSN: 0924-3275.
[70] D. Mullensiefen, K. Frieler, Optimizing measures of melodic similarity of folk songs, in: Proceedings of Fifth International Conference on Music Information Retrieval, Barcelona, Spain, 10–15 October 2004, pp. 274–280.
[71] J. Garbers, et al., Using pitch stability among a group of aligned query melodies to retrieve unidentified variant melodies, in: Proceedings of Eighth International Conference on Music Information Retrieval, Vienna, Austria, 23–30 September 2007, pp. 451–456.
[72] A. Volk, et al., A manual annotation method for melodic similarity and the study of melody feature sets, in: Proceedings of Ninth International Conference on Music Information Retrieval, Philadelphia, USA, 14–18 September 2008, pp. 101–106.
[73] P. Toivianen, T. Eerola, Classification of musical metre with autocorrelation and discriminant functions, in: Proceedings of Sixth International Conference on Music Information Retrieval, London, UK, 11–15 September 2005, pp. 351–357.
[74] A. Volk, et al., Applying rhythmic similarity based on inner metric analysis to folk music, in: Proceedings of Eighth International Conference on Music Information Retrieval, Vienna, Austria, 23–30 September 2007, pp. 293–296.
[75] G.T. Toussaint, A comparison of rhythmic similarity measures, in: Proceedings of Fifth International Conference on Music Information Retrieval, Barcelona, Spain, 10–15 October 2004, pp. 242–245.
[76] K. Gustafson, The graphical representation of rhythm, in: (PROPH) Progress Reports from Oxford Phonetics, vol. 3, University of Oxford, 1988, pp. 6–26.
[77] K. Frieler, Visualizing music on the metrical circle, in: Proceedings of Eighth International Conference on Music Information Retrieval, Vienna, Austria, 23–30 September 2007, pp. 291–292.
[78] L.A. Uitdenbogerd, I.H. Suyoto, Microtonal matching with MTRI., in: W.B. Hewlett, E. Selfridge-Field (Eds.), Computing in Musicology, vol. 14, MIT Press, Cambridge, MA, 2006, pp. 334–341.
[79] Audio-file transfer and exchange—file format for transferring digital audio data between systems of different type and manufacture, AES31-2-2006: AES standard on network and file transfer of audio, 2006.
[80] L.-C. Koch, et al., The Berlin phonogramm-archiv: a treasury of sound recordings, Acoustical Science and Technology 25 (4) (2004) 227–231.
[81] N. Corthaut, et al., Connecting the dots: music metadata generation, in: Proceedings of Ninth International Conference on Music Information Retrieval, Philadelphia, USA, 14–18 September 2008, pp. 249–254.
[82] C. Lai, et al., Metadata infrastructure for sound recordings, in: Proceedings of Eighth International Conference on Music Information Retrieval, Vienna, Austria, 23–30 September 2007, pp. 157–158.
[83] C. Lai, I. Fujinaga, Metadata data dictionary for analogue sound recordings, in: Proceedings of the Joint Conference on Digital Libraries. Chapel Hill, NC, 2006, p. 344.
[84] O. Cornelis, et al., Development and Possibilities of a Digital Database for Ethnic Musical Metadata; From obvious to oblivion, in: Proceedings of the Fifth incontro biennale internazionale sul restauro audio, Treviso, Italy, 6-7 October 2006, in press.
[85] G. De Tré, et al., Aggregating constraint satisfaction degrees expressed by possibilistic truth values, IEEE Transactions on Fuzzy Systems 11 (3) (2003) 361–368.
[86] G. De Tré, et al., Ranking the possible alternatives in flexible querying: an extended possibilistic approach, in: Lecture Notes In Computer Science, vol. 2869, 2003, pp. 204–211.
[87] M. Fingerhut, Real music libraries in the virtual future: for an integrated view of music and music information, in: M. Leman, O. Cornelis (Eds.), Digitale Bibliotheken voor muzikale audio, KVAB, Belgium, 2005, pp. 73–81.
[88] R. Leydi, L'altre Musica, (Ed.), Giunti Ricordi, Milan, 1991.

[89] G. Adler, Unfang, Vierteljahrschrift für Musikwissenschaft 1, 1895.
[90] A. Seeger, Dealing with the ethical and legal constraints of information access, in: Proceedings of Fourth International Conference on Music Information Retrieval, Baltimore, Maryland, USA, 26–30 October 2003.
[91] C. Landone, et al., Enabling access to sound archives through integration, enrichment and retrieval: the EASAIER Project, in: Proceedings of Eighth International Conference on Music Information Retrieval, Vienna, Austria, 23–30 September 2007, pp. 159–160.
[92] J. Futrelle, J.S. Downie, Interdisciplinary communities and research issues in music information retrieval, in: Proceedings of Third International Conference on Music Information Retrieval, Paris, France, 13–17 October 2002, pp. 215–221.
[93] J.S. Downie, Toward the scientific evaluation of music information retrieval systems, in: Proceedings of Fourth International Conference on Music Information Retrieval, Baltimore, Maryland, USA, 26–30 October 2003, pp. 25–32.
[94] W.A. Sethares, Tuning, Timbre, Spectrum, Scale, Springer, London, New York, 1998.
[95] O. Cornelis, et al., Development and Possibilities of a Digital Database for Ethnic Musical Metadata; From obvious to oblivion, in: Proceedings of the Fifth incontro biennale internazionale sul restauro audio, Treviso, Italy, 6-7 October 2006, in press.

Six, J., Cornelis, O., & Leman, M. (2013). Tarsos, a Modular Platform for precise Pitch Analysis of Western and non-Western music. Journal of New Music Research. 42 (2) pp. 113-129.

Journal of New Music Research, 2013
Vol. 42, No. 2, 113–129, http://dx.doi.org/10.1080/09298215.2013.797999

Routledge
Taylor & Francis Group

Tarsos, a Modular Platform for Precise Pitch Analysis of Western and Non-Western Music

Joren Six[1] , Olmo Cornelis[1] and Marc Leman[2]

[1]*University College Ghent, Belgium;* [2]*Ghent University, Belgium*

Abstract

This paper presents Tarsos, a modular software platform used to extract and analyse pitch organization in music. With Tarsos pitch estimations are generated from an audio signal and those estimations are processed in order to form musicologically meaningful representations. Tarsos aims to offer a flexible system for pitch analysis through the combination of an interactive user interface, several pitch estimation algorithms, filtering options, immediate auditory feedback and data output modalities for every step. To study the most frequently used pitches, a fine-grained histogram that allows up to 1200 values per octave is constructed. This allows Tarsos to analyse deviations in Western music, or to analyse specific tone scales that differ from the 12 tone equal temperament, common in many non-Western musics. Tarsos has a graphical user interface or can be launched using an API—as a batch script. Therefore, it is fit for both the analysis of individual songs and the analysis of large music corpora. The interface allows several visual representations, and can indicate the scale of the piece under analysis. The extracted scale can be used immediately to tune a MIDI keyboard that can be played in the discovered scale. These features make Tarsos an interesting tool that can be used for musicological analysis, teaching and even artistic productions.

1. Introduction

In the past decennium, several computational tools became available for extracting pitch from audio recordings (Clarisse et al., 2002; de Cheveigné & Hideki, 2002; Klapuri, 2003). Pitch extraction tools are prominently used in a wide range of studies that deal with analysis, perception and retrieval of music. However, up to recently, less attention has been paid to tools that deal with distributions of pitch in music.

The present paper presents a tool, called Tarsos, that integrates existing pitch extraction tools in a platform that allows the analysis of pitch distributions. Such pitch distributions contain a lot of information, and can be linked to tunings, scales, and other properties of musical performance. The tuning is typically reflected in the distance between pitch classes. Properties of musical performance may relate to pitch drift within a single piece, or to influence of enculturation (as is the case in African music culture, see Moelants, Cornelis and Leman (2009)). A major feature of Tarsos is concerned with processing audio-extracted pitches into pitch and pitch class distributions from which further properties can be derived.

Tarsos provides a modular platform used for pitch analysis—based on pitch extraction from audio and pitch distribution analysis—with a flexibility that includes:

- The possibility to focus on a part of a song by selecting graphically displayed pitch estimations in the melograph.
- A zoom function that allows focusing on global or detailed properties of the pitch distribution.
- Real-time auditory feedback. A tuned MIDI synthesizer can be used to hear pitch intervals.
- Several filtering options to get clearer pitch distributions or a more discretized melograph, which helps during transcription.

In addition, a change in one of the user interface elements is immediately propgated through the whole processing chain, so that pitch analysis becomes easy, adjustable and verifiable.

This paper is structured as follows. First, we present a general overview of the different processing stages of Tarsos, beginning with the low level audio signal stage and ending with pitch distributions and their musicological meaning. In the next part, we focus on some case studies and give a scripting example. The next part elaborates on the musical aspects of Tarsos and refers to future work. The fifth and final part of the main text contains a conclusion.

Correspondence: Joren Six, University College Ghent, School of Arts, Hoogpoort 64, Gent, 9000 Belgium. E-mail: joren.six@hogent.be

2. The Tarsos platform

Figure 1 shows the general flow of information within Tarsos. It starts with an audio file as input. The selection of a pitch estimation algorithm leads to pitch estimations, which can be represented in different ways. This representation can be further optimized, using different types of filters for peak selection. Finally, it is possible to produce an audio output of the obtained results. Based on that output, the analysis–representation–optimization cycle can be refined. All steps contain data that can be exported in different formats. The obtained pitch distribution and scale itself can be saved as a scala file which in turn can be used as input, overlaying the estimation of another audio file for comparison.

In what follows, we go deeper into several processing aspects, dependencies, and particularities. In this section we first discuss how to extract pitch estimations from audio. We illustrate how these pitch estimations are visualized within Tarsos. The graphical user interface is discussed. The real-time and output capabilities are described, and this section ends with an explanation about scripting for the Tarsos API. As a reminder: there is a manual available for Tarsos at http://tarsos.0110.be/tag/JNMR.

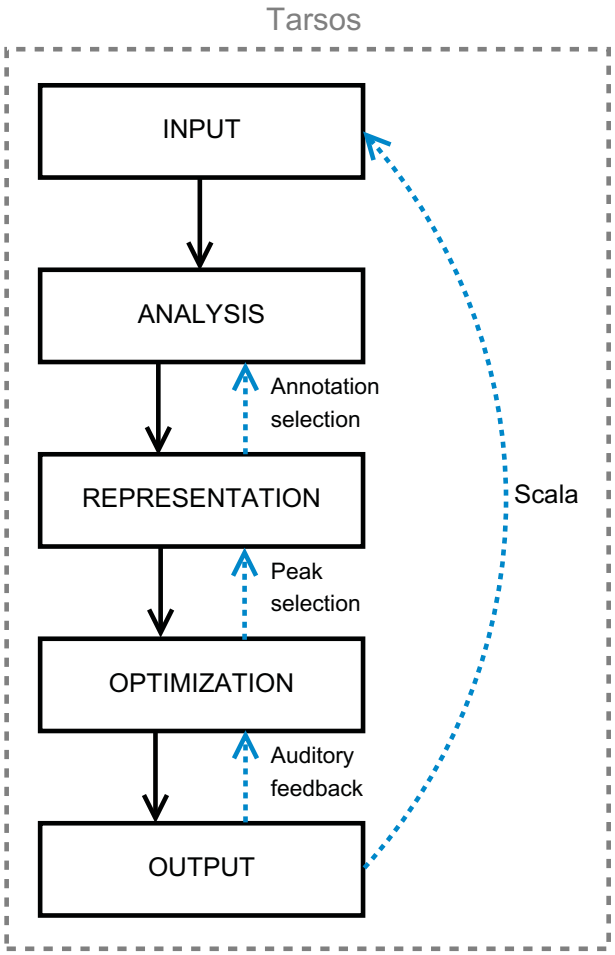


Fig. 1. The main flow of information within Tarsos.

2.1 Extracting pitch estimations from audio

Prior to the step of pitch estimation, one should take into consideration that in certain cases, audio preprocessing can improve the subsequent analysis within Tarsos. Depending on the source material and on the research question, preprocessing steps could include noise reduction, band-pass filtering, or harmonic/percussive separation Nobutaka et al. (2010). Audio preprocessing should be done outside of the Tarsos tool. The, optionally preprocessed, audio is then fed into Tarsos and converted to a standardized format.[1]

The next step is to generate pitch estimations. Each selected block of audio file is examined and pitches are extracted from it. In Figure 2, this step is located between the input and the signal block phases. Tarsos can be used with external and internal pitch estimators. Currently, there is support for the polyphonic MAMI pitch estimator (Clarisse et al., 2002) and any VAMP plug-in (Cannam, Landone, & Sandler, 2010) that generates pitch estimations. The external pitch estimators are platform dependent and some configuration needs to be done to get them working. For practical purposes, platform independent implementations of two pitch detection algorithms are included, namely, YIN (de Cheveigné & Hideki, 2002) and MPM (McLeod & Wyvill, 2005). They are available without any configuration. Thanks to a modular design, internal and external pitch detectors can be easily added. Once correctly configured, the use of these pitch modules is completely transparent, as extracted pitch estimations are transformed to a unified format, cached, and then used for further analysis at the symbolic level.

2.2 Visualizations of pitch estimations

Once the pitch detection has been performed, pitch estimations are available for further study. Several types of visualizations can be created, which lead, step by step, from pitch estimations to pitch distribution and scale representation. In all these ca graphs the *cent* unit is used. The cent divides each octave into 1200 equal parts. In order to use the cent unit for determining absolute pitch, a reference frequency of 8.176 Hz has been defined,[2] which means that 8.176 Hz equals 0 cents, 16.352 Hz equals 1200 cents and so on.

A first type of visualization is the melograph representation, which is shown in Figure 3. In this representation, each estimated pitch is plotted over time. As can be observed, the pitches are not uniformly distributed over the pitch space, and form a clustering around 5883 cents.

A second type of visualization is the pitch histogram, which shows the pitch distribution regardless of time. The pitch histogram is constructed by assigning each pitch estimation

[1] The conversion is done using FFMPEG, a cross-platform command line tool to convert multimedia files between formats. The default format is PCM WAV with 16 bits per sample, signed, little endian, 44.1 kHz. Furthermore, all channels are down mixed to mono.

[2] See Appendix B for a discussion about pitch representation in cents and the seemingly arbitrary reference frequency of 8.176 Hz.

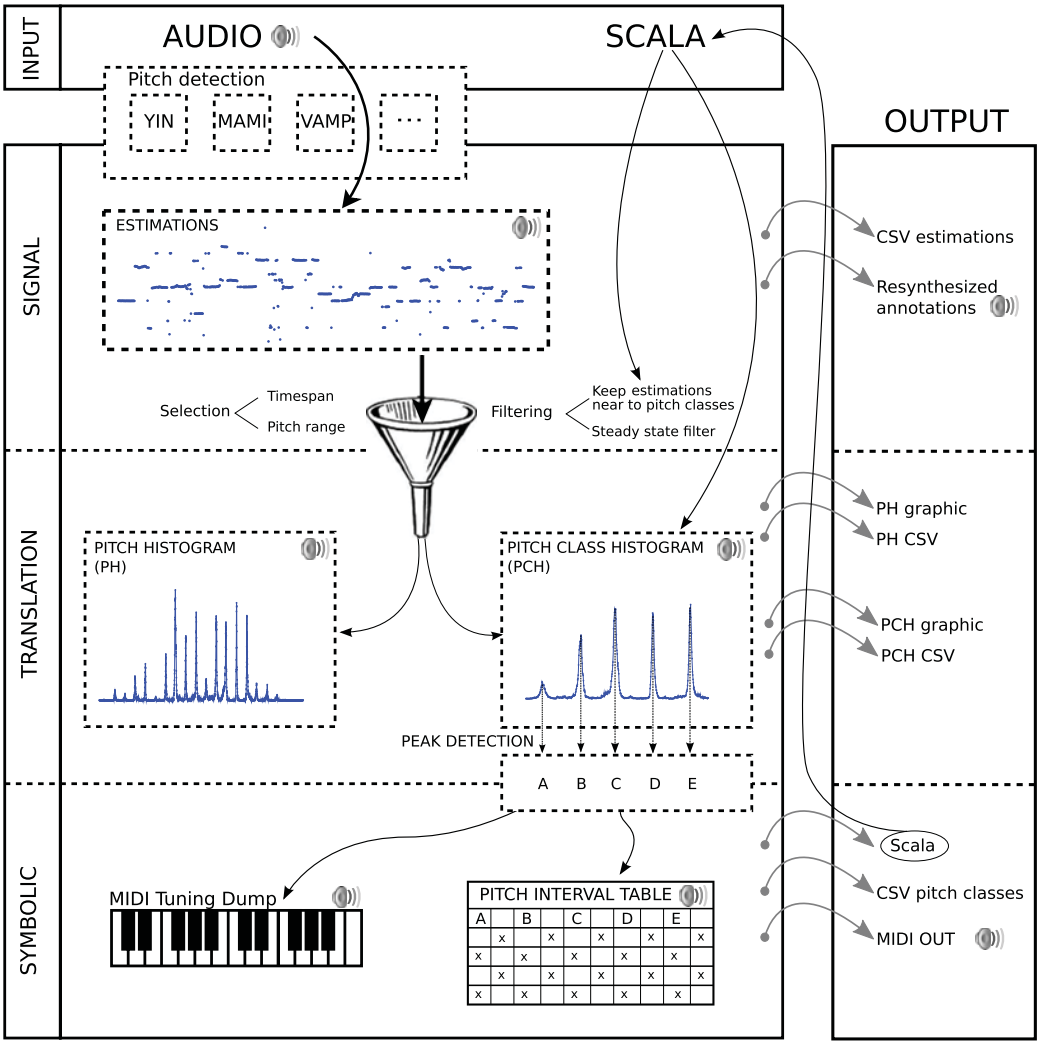


Fig. 2. Detailed block diagram representing all components of Tarsos, from input to output, from signal level to symbolic level. All additional features (selection, filtering, listening) are visualized (where they come into play). Each step is described in more detail in Section 3.

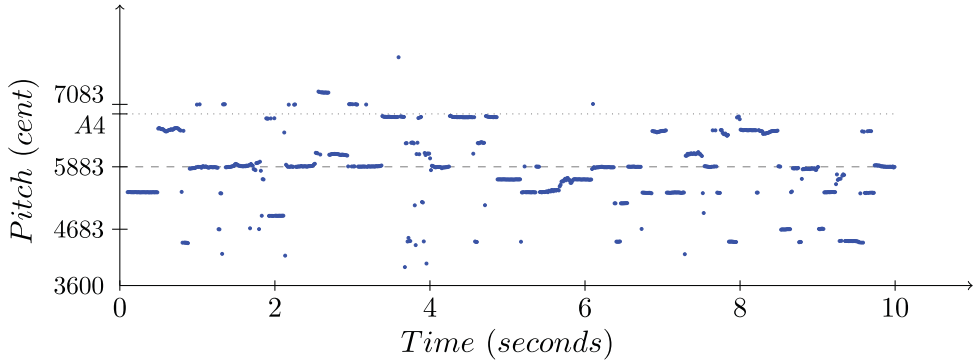


Fig. 3. A melograph representation. Estimations of the first ten seconds of an Indonesian Slendro piece are shown. It is clear that pitch information is horizontally clustered, e.g. the cluster around 5883 cents, indicated by the dashed horizontal line. For reference a dotted horizontal line with A4, 440Hz is also present.

in time to a bin between 0 and 14,400 cents,[3] spanning octaves. As shown in Figure 4, the peak at 5883 cents is now 12 clearly visible. The height of a peak represents the total number of times a particular pitch is estimated in the selected audio. The pitch range is the difference between the highest and lowest pitch. The graph further reveals that some peaks appear every 1200 cents, or every octave.

[3] 14,400 absolute cents is equal to 33,488 Hz, well above human hearing.

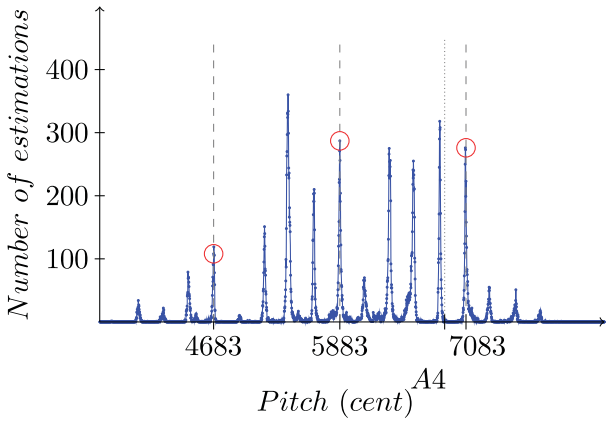


Fig. 4. A pitch histogram with an Indonesian Slendro scale. The circles mark the most estimated pitch classes. The dashed vertical lines show the same pitch class in different octaves. A dotted vertical line with A4, 440Hz, is used as a reference for the diapason.

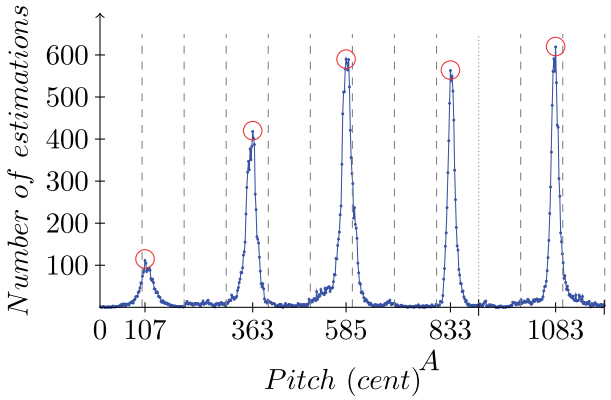


Fig. 5. A pitch class histogram with an Indonesian Slendro scale. The circles mark different pitch classes. For reference, the dashed lines represent the Western equal temperament. The pitch class A is marked with a dotted line.

A third type of visualization is the pitch *class* histogram, which is obtained by adding each bin from the pitch histogram to a corresponding modulo 1200 bin. Such a histogram reduces the pitch distribution to one single octave. A peak thus represents the total duration of a pitch *class* in a selected block of audio. Notice that the peak at 5883 cents in the pitch histogram (Figure 4) now corresponds to the peak at 1083 cents in the pitch class histogram (Figure 5).

It can also be useful to select only filter pitch estimations that make up the pitch class histogram. The most obvious 'filter' is to select only an interesting timespan and pitch range. The distributions can be further manipulated using other filters and peak detection. The following three filters are implemented in Tarsos.

The first is an *estimation quality* filter. It simply removes pitch estimations from the distribution below a certain quality threshold. Using YIN, the quality of an estimation is related to the periodicity of the block of sound analysed. Keeping only high quality estimations should yield clearer pitch distributions.

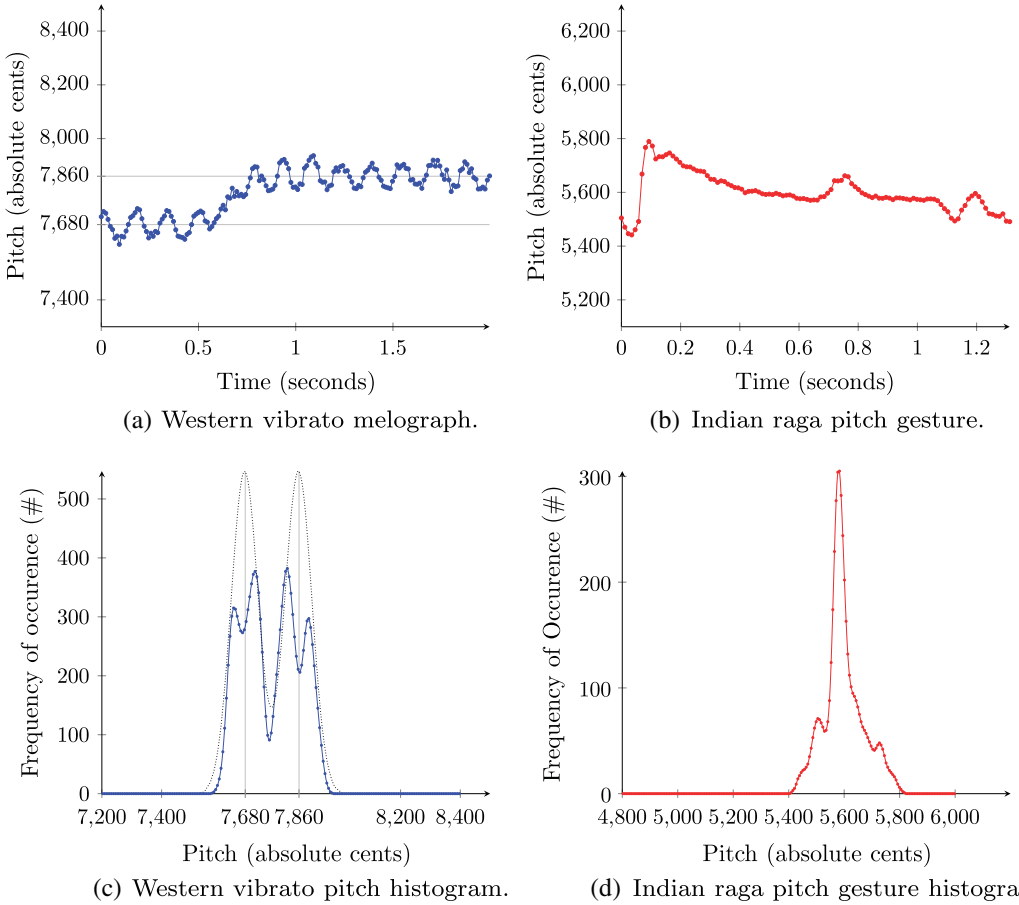
The second is called a *near to pitch class filter*. This filter only allows pitch estimations which are close to previously identified pitch classes. The pitch range parameter (in cents) defines how much ornamentations can deviate from the pitch classes. Depending on the music and the research question, one needs to be careful with this—and other—filters. For example, a vibrato makes pitch go up and down—pitch modulation—and is centred around a pitch class. Figure 6(a) gives an example of Western vibrato singing. The melograph reveals the ornamental singing style, based on two distinct pitch classes. The two pitch classes are hard to identify with the histogram 6(c) but are perceptually there, they are made clear with the dotted gray line. In contrast, Figure 6(b) depicts a more continuous glissando which is used as a building block to construct a melody in an Indian raga. For these cases, Krishnaswamy (2004b) introduced the concept of two-dimensional 'melodic atoms'. (In Henbing and Leman (2007) it is shown how elementary bodily gestures are related to pitch and pitch gestures.) The histogram of the pitch gesture, Figure 6(d), suggests one pitch class while a fundamentally different concept of tone is used. Applying the near to pitch class filter on this type of music could result in incorrect results. The goal of this filter is to get a clearer view on the melodic contour by removing pitches between pitch classes, and to get a clearer pitch class histogram.

The third filter is a *steady state filter*. The steady state filter has a time and pitch range parameter. The filter keeps only consecutive estimations that stay within a pitch range for a defined number of milliseconds. The default values are 100 ms within a range of 15 cents. The idea behind it is that only 'notes' are kept and transition errors, octave errors and other short events are removed.

Once a selection of the estimations are made or, optionally, other filters are used, the distribution is ready for peak detection. The peak detection algorithm looks for each position where the derivative of the histogram is zero, and a local height score is calculated with the formula in (1). The local height score s_w is defined for a certain window w, μ_w is the average height in the window, σ_w refers to the standard deviation of the height in the window. The peaks are ordered by their score and iterated, starting from the peak with the highest score. If peaks are found within the window of the current peak, they are removed. Peaks with a local height score lower than a defined threshold are ignored. Since we are looking for pitch classes, the window w wraps around the edges: there is a difference of 20 cents between 1190 cent and 10 cents.

$$s_w = \frac{height - \mu_w}{\sigma_w}. \quad (1)$$

Figure 7 shows the local height score function applied to the pitch class histogram shown in Figure 5. The desired levelling effect of the local height score is clear, as the small peak at 107 cents becomes much more defined. The threshold is also shown. In this case, it eliminates the noise at around 250 cents. The noise is caused by the small window size and local height deviations, but it is ignored by setting threshold t. The per-



(a) Western vibrato melograph. (b) Indian raga pitch gesture.

(c) Western vibrato pitch histogram. (d) Indian raga pitch gesture histogram.

Fig. 6. Visualization of pitch contours of Western and Indian singing; notice the fundamentally different concept of tone. In the Western example two distinct pitches are used, they are made clear with the dotted gray lines. In (c) two dotted gray curves are added, they represent the two perceived pitches.

formance of the peak detection depends on two parameters, namely, the window size and the threshold. Automatic analysis either uses a general pre-set for the parameters or tries to find the most stable setting with an exhaustive search. Optionally gaussian smoothing can be applied to the pitch class histogram, which makes peak detection more straightforward. Manual intervention is sometimes needed, by fiddling with the two parameters a user can quickly browse through several peak detection result candidates.

Once the pitch classes are identified, a pitch class interval matrix can be constructed. This is the fourth type of representation, which is shown in Table 1. The pitch class interval matrix represents the found pitch classes, and shows the intervals between the pitch classes. In our example, a perfect fourth,[4] a frequency ratio of 4/3 or 498 cent, is present between pitch class 585 and 1083. This means that a perfect fifth, a frequency ratio of $\frac{2/1}{4/3} = 3/2$ or $1200 - 498 = 702$ cent, is also present.[5]

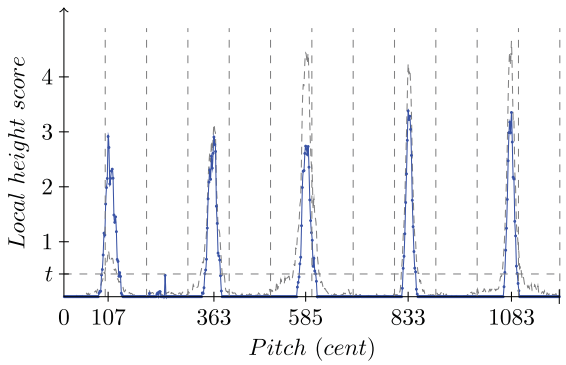


Fig. 7. A local height score function used to detect peaks in a pitch class histogram. Comparing the original histogram of Figure 5 with the local height score shows the levelling effect of the local height score function. The dashed vertical lines represent the Western equal temperament, the dashed horizontal line the threshold t.

2.3 The interface

Most of the capabilities of Tarsos are used through the graphical user interface (Figure 8). The interface provides a way to explore pitch organization within a musical piece. However,

[4]The perfect fourth and other musical intervals are here used in their physical meaning. The physical perfect fourth is sometimes called just fourth, or perfect fourth in just intonation.

[5]See Appendix B to see how ratios translate to cent values.

Table 1. Pitch classes (P.C.) and pitch class intervals, both in cents. The same pentatonic Indonesian slendro is used as in Figure 5. A perfect fifth and its dual, a perfect fourth, are marked by a bold font.

P.C.	107	364	585	833	1083
107	0	256	478	726	976
364	944	0	221	470	719
585	722	979	0	248	**498**
833	474	730	952	0	250
1083	224	481	**702**	950	0

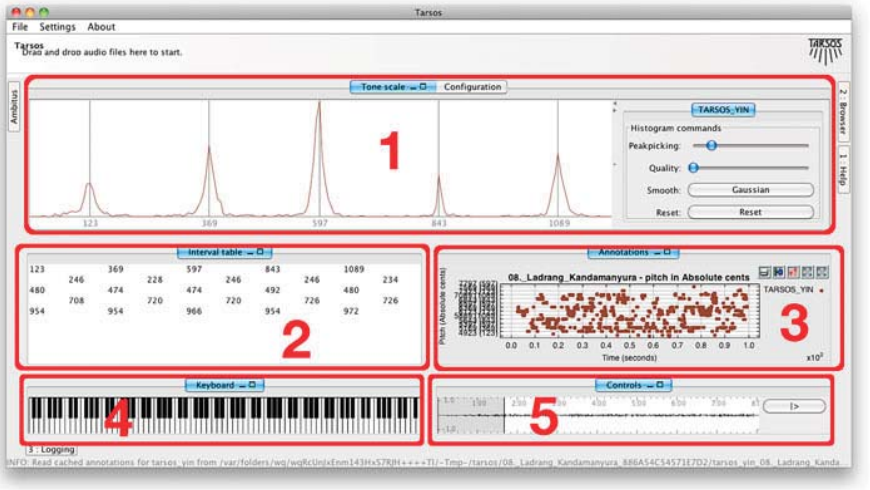


Fig. 8. A screenshot of Tarsos with 1 a pitch class histogram, 2 a pitch class interval table, 3 a melograph with pitch estimations, 4 a midi keyboard and 5 a waveform.

the main flow of the process, as described above, is not always as straightforward as the example might suggest. More particularly, in many cases of music from oral traditions, the peaks in the pitch class histogram are not always well-defined (see Section 4). Therefore, the automated peak detection may need manual inspection and further manual fine-tuning in order to correctly identify a songs' pitch organization. The user interface was designed specifically for having a flexible environment where all windows with representations communicate their data. Tarsos has the attractive feature that all actions, like the filtering actions mentioned in Section 2.2, are updated for each window in real-time.

One way to closely inspect pitch distributions is to select only a part of the estimations. In the block diagram of Figure 2, this is represented by the funnel. Selection in time is possible using the waveform view (Figure 8–5). For example, the aim could be a comparison of pitch distributions at the beginning and the end of a piece, to reveal whether a choir lowered or raised its pitch during a performance (see Section 4 for a more elaborate example).

Selection in pitch range is possible and can be combined with a selection in time using the melograph (Figure 8–3). One may select the melodic range so as to exclude pitched percussion, and this could yield a completely different pitch class histogram. This feature is practical, for example when a flute melody is accompanied with a low-pitched drum and when you are only interested in flute tuning. With the melograph it is also possible to zoom in on one or two notes, which is interesting for studying pitch contours. As mentioned earlier, not all music is organized by fixed pitch classes. An example of such pitch organization is given in Figure 6(b), a fragment of Indian music where the estimations contain information that cannot be reduced to fixed pitch classes.

To allow efficient selection of estimations in the time and frequency, they are stored in a kd-tree (Bentley, 1975). Once such a selection of estimations is made, a new pitch histogram is constructed and the pitch class histogram view (Figure 8–1) changes instantly.

Once a pitch class histogram is obtained, peak detection is a logical next step. With the user interface, manual adjustment of the automatically identified peaks is possible. New peak locations can be added and existing ones can be moved or deleted. In order to verify the pitch classes manually, it is possible to click anywhere on the pitch class histogram. This sends a MIDI-message with a pitch bend to synthesize a sound with a pitch that corresponds to the clicked location. Changes made to the peak locations propagate instantly throughout the interface.

The pitch class interval matrix (Figure 8–2) shows all new pitch class intervals. Reference pitches are added to the melograph and MIDI tuning messages are sent (see Section 2.5). The pitch class interval matrix is also interactive. When an

interval is clicked on, the two pitch classes that create the interval sound at the same time. The dynamics of the process and the combination of both visual and auditory clues makes manually adjusted, precise peak extraction, and therefore tone scale detection, possible. Finally, the graphical display of a piano keyboard in Tarsos allows us to play in the (new) scale. This feature can be executed on a computer keyboard as well, where notes are projected on keys. Any of the standard MIDI instrument sounds can be chosen.

It is possible to shift the pitch class histogram up—or downwards. The data is then viewed as a repetitive, octave based, circular representation. In order to compare scales, it is possible to upload a previously detected scale (see Section 2.5) and shift it, to find a particular fit. This can be done by hand, exploring all possibilities of overlaying intervals, or the best fit can be suggested by Tarsos.

2.4 Real-time capabilities

Tarsos is capable of real-time pitch analysis. Sound from a microphone can be analysed and immediate feedback can be given on the played or sung pitch. This feature offers some interesting new use-cases in education, composition, and ethnomusicology.

For educational purposes, Tarsos can be used to practice singing quarter tones. Not only the real time audio is analysed, but also an uploaded scale or previously analysed file can be listened to by clicking on the interval table or by using the keyboard. Singers or string players could use this feature to improve their intonation regardless of the scale they try to reach.

For compositional purposes, Tarsos can be used to experiment with microtonality. The peak detection and manual adjustment of pitch histograms allows the construction of any possible scale, with the possibility of setting immediate harmonic and melodic auditory feedback. Use of the interval table and the keyboard, make experiments in interval tension and scale characteristics possible. Musicians can tune (ethnic) instruments according to specific scales using the direct feedback of the real-time analysis. Because of the MIDI messages, it is also possible to play the keyboard in the same scale as the instruments at hand.

In ethnomusicology, Tarsos can be a practical tool for direct pitch analysis of various instruments. Given the fact that pitch analysis results show up immediately, microphone positions during field recordings can be adjusted on the spot to optimize measurements.

2.5 Output capabilities

Tarsos contains export capabilities for each step, from the raw pitch estimations until the pitch class interval matrix. The built-in functions can export the data as comma separated text files, charts, TEX-files, and there is a way to synthesize estimations. Since Tarsos is scriptable there is also a possibility to add other export functions or modify the existing functions. The API and scripting capabilities are documented on the Tarsos website: http://tarsos.0110.be/tag/JNMR.

For pitch class data, there is a special standardized text file defined by the Scala[6] program. The Scala file format has the `.scl` extension. The Scala program comes with a dataset of over 3900 scales ranging from historical harpsichord temperaments over ethnic scales to scales used in contemporary music. Recently this dataset has been used to find the universal properties of scales (Honingh & Bod, 2011). Since Tarsos can export scala files it is possible to see if the star-convex structures discussed in Honingh and Bod (2011) can be found in scales extracted from real audio. Tarsos can also parse Scala files, so that comparison of theoretical scales with tuning practice is possible. This feature is visualized by the upwards Scala arrow in Figure 2. When a scale is overlaid on a pitch class histogram, Tarsos finds the best fit between the histogram and the scala file.

2.6 Scripting capabilities

Processing many audio files with the graphical user interface quickly becomes tedious. Scripts written for the Tarsos API can automate tasks and offer a possibility to utilize Tarsos' building blocks in entirely new ways. Tarsos is written in Java, and is extendable using scripts in any language that targets the JVM (Java Virtual Machine) like JRuby, Scala[7] and Groovy. For its concurrency support, concise syntax and seamless interoperability with Java, the Scala programming languages are used in example scripts, although the concepts apply to scripts in any language. The number of applications for the API is only limited by the creativity of the developer using it. Tasks that can be implemented using the Tarsos API are for example:

> *Tone scale recognition:* given a large number of songs and a number of tone scales in which each song can be brought, guess the tone scale used for each song. In Section 3.4 this task is explained in detail and effectively implemented.
> *Modulation detection:* this task tries to find the moments in a piece of music where the pitch class histogram changes from one stable state to another. For Western music this could indicate a change of mode, a modulation. This task is similar as the one described in Mearns, Benetos and Dixon (2011). With the Tarsos API you can compare windowed pitch histograms and detect modulation boundries.
> *Evolutions in tone scale use:* this task tries to find evolutions in tone scale use in a large number of songs from a certain region over a long period of time. Are some pitch intervals becoming more popular than others? In Moelants et al. (2009) this is done for a set of African songs.

[6] See http://www.huygens-fokker.org/scala/

[7] Please do not confuse the general purpose Scala programming language with the tool to experiment with tunings, the Scala program.

Acoustic Fingerprinting: it is theorized in Tzanetakis, Ermolinskyi and Cook (2002) that pitch class histograms can serve as an acoustic fingerprint for a song. With the building blocks of Tarsos: pitch detection, pitch class histogram creation and comparison this was put to the test by Six and Cornelis (2012).

The article by Tzanetakis et al. (2002) gives a good overview of what can be done using pitch histograms and, by extension, the Tarsos API. To conclude: the Tarsos API enables developers to quickly test ideas, execute experiments on large sets of music and leverage the features of Tarsos in new and creative ways.

3. Exploring Tarsos' capabilities through case studies

In what follows, we explore Tarsos' capabilities using case studies in non-Western music. The goal is to focus on problematic issues such as the use of different pitch extractors, music with pitch drift, and last but not least, the analysis of large databases.

3.1 Analysing a pitch histogram

We will first consider the analysis of a song that was recorded in 1954 by missionary Scohy-Stroobants in Burundi. The song is performed by a singing soloist, Lonard Ndengabaganizi. The recording was analysed with the YIN pitch detection method and a pitch class histogram was calculated: it can be seen in Figure 9. After peak detection on this histogram, the following pitch intervals were detected: 168, 318, 168, 210, and 336 cents. The detected peaks and all intervals are shown in an interval matrix (see Figure 9). It can be observed that this is a pentatonic division that comprises small and large intervals, which is different from an equal tempered or meantone division. Interestingly, the two largest peaks define a fifth interval, which is made of a pure minor third (318 cents) and a pure major third (378 cents) that lies between the intervals 168+210 = 378 cents). In addition, a mirrored set of intervals is present, based on 168-318-168 cents. This phenomena is also illustrated by Figure 9.

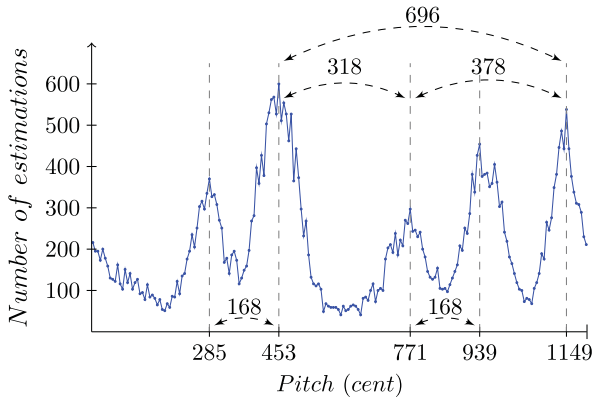


Fig. 9. This song uses an unequally divided pentatonic tone scale with mirrored intervals 168-318-168, indicated on the pitch class histogram. Also indicated is a near perfect fifth consisting of a pure minor and pure major third.

3.2 Different pitch extractors

However, Tarsos has the capability to use different pitch extractors. Here we show the difference between seven pitch extractors on a histogram level. A detailed evaluation of each algorithm cannot be covered in this article but can be found in the cited papers. The different pitch extractors are:

- YIN (de Cheveigné & Hideki, 2002) (YIN) and the McLeod Pitch Method (MPM), which is described in McLeod, & Wyvill (2005), are two time-domain pitch extractors. Tarsos contains a platform independent implementation of the algorithms.
- Spectral Comb (SC), Schmitt trigger (Schmitt) and Fast Harmonic Comb (FHC) are described in Brossier (2006). They are available for Tarsos through VAMP-plugins (Cannam, 2008);
- MAMI 1 and MAMI 6 are two versions of the same pitch tracker. MAMI 1 only uses the most present pitch at a certain time, MAMI 6 takes the six most salient pitches at a certain time into account. The pitch tracker is described in Clarisse et al. (2002).

Figure 10 shows the pitch histogram of the same song as in the previous section, which is sung by an unaccompanied young man. The pitch histogram shows a small tessitura and wide pitch classes. However, the general contour of the histogram is more or less the same for each pitch extraction method, five pitch classes can be distinguished in about one-and-a-half octaves, ranging from 5083 to 6768 cents. Two methods stand out. Firstly, MAMI 6 detects pitch in the lower and higher regions. This is due to the fact that MAMI 6 always gives six pitch estimations in each measurement sample. In this monophonic song this results in octave—halving and doubling—errors and overtones. Secondly, the Schmitt method also stands out because it detects pitch in regions where other methods detect a lot less pitches, e.g. between 5935 and 6283 cents.

Figure 11 shows the pitch class histogram for the same song as in Figure 10, now collapsed into one octave. It clearly shows that it is hard to determine the exact location of each pitch class. However, all histogram contours look similar except for the Schmitt method, which results in much less well-defined peaks. The following evaluation shows that this is not the only case.

In order to be able to gain some insight into the differences between the pitch class histograms resulting from different pitch detection methods, the following procedure was used: for each song in a data set of more than 2800 songs—a random selection of the music collection of the Belgian Royal

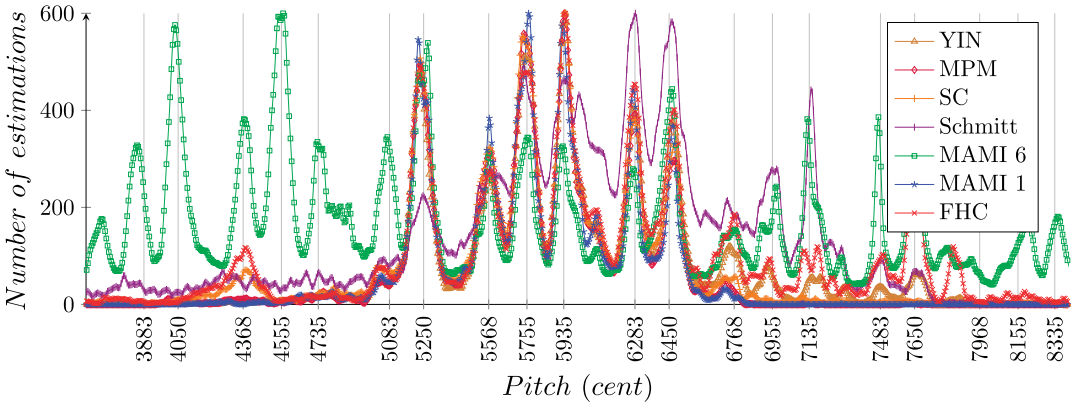


Fig. 10. Seven different pitch histograms of a traditional Rwandese song. Five pitch classes repeat every octave. The Schmitt trigger (Schmitt) results in much less well-defined peaks in the pitch histogram. MAMI 6 detects much more information to be found in the lower and higher regions, this is due to the fact that it always gives six pitch estimations, even if they are octave errors or overtones.

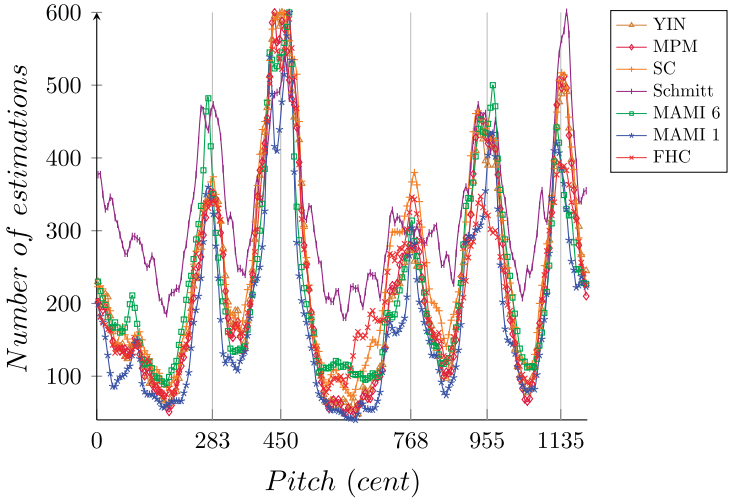


Fig. 11. Seven different pitch class histograms of a traditional Rwandese song. Five pitch classes can be distinguished but it is clear that it is hard to determine the exact location of each pitch class. The Schmitt trigger (Schmitt) results in a lot fewer well-defined peaks in the pitch class histogram.

Museum of Central Africa (RMCA)—seven pitch class histograms were created by the pitch detection methods. The overlap—a number between zero and one—between each pitch class histogram pair was calculated. A sum of the overlap between each pair was made and finally divided by the number of songs. The resulting data can be found in Table 2. Here histogram overlap or intersection is used as a distance measure because Gedik and Bozkurt (2010) show that this measure works best for pitch class histogram retrieval tasks. The overlap $c(h_1, h_2)$ between two histograms h_1 and h_2 with K classes is calculated with Equation (2). For an overview of alternative correlation measures between probability density functions see Cha (2007).

$$c(h_1, h_2) = \frac{\sum_{k=0}^{K-1} \min(h_1(k), h_2(k))}{\max\left(\sum_{k=0}^{K-1} h_1(k), \sum_{k=0}^{K-1} h_2(k)\right)}. \quad (2)$$

Table 2 shows that there is, on average, a large overlap of 81%, between the pitch class histograms created by YIN and those by MPM. This can be explained by the fact that the two pitch extraction algorithms are very much alike: both operate in the time-domain with autocorrelation. The table also shows that Schmitt generates rather unique pitch class histograms. On average there is only 55% overlap with the other pitch class histogram. This performance was already expected during the analysis of one song (above).

The choice for a particular pitch detection method depends on the music and the analysis goals. The music can be monophonic, homophonic or polyphonic, different instrumentation and recording quality all have influence on pitch estimators. Users of Tarsos are encouraged to try out which pitch detection method suits their needs best. Tarsos' scripting API—see Section 3.4—can be helpful when optimizing combinations of pitch detection methods and parameters for an experiment.

Table 2. Similarity matrix showing the overlap of pitch class histograms for seven pitch detection methods. The similarities are the mean of 2484 audio files. The last row shows the average of the overlap for a pitch detection method.

	YIN	MPM	Schmitt	FHC	SC	MAMI 1	MAMI 6
YIN	1.00	**0.81**	*0.41*	0.65	0.62	0.69	0.61
MPM	**0.81**	1.00	0.43	0.67	0.64	0.71	0.63
Schmitt	*0.41*	0.43	1.00	0.47	0.53	0.42	0.56
FHC	0.65	0.67	0.47	1.00	0.79	0.67	0.66
SC	0.62	0.64	0.53	0.79	1.00	0.65	0.70
MAMI 1	0.69	0.71	0.42	0.67	0.65	1.00	0.68
MAMI 6	0.61	0.63	0.56	0.66	0.70	0.68	1.00
Average	0.69	0.70	*0.55*	0.70	0.70	0.69	0.69

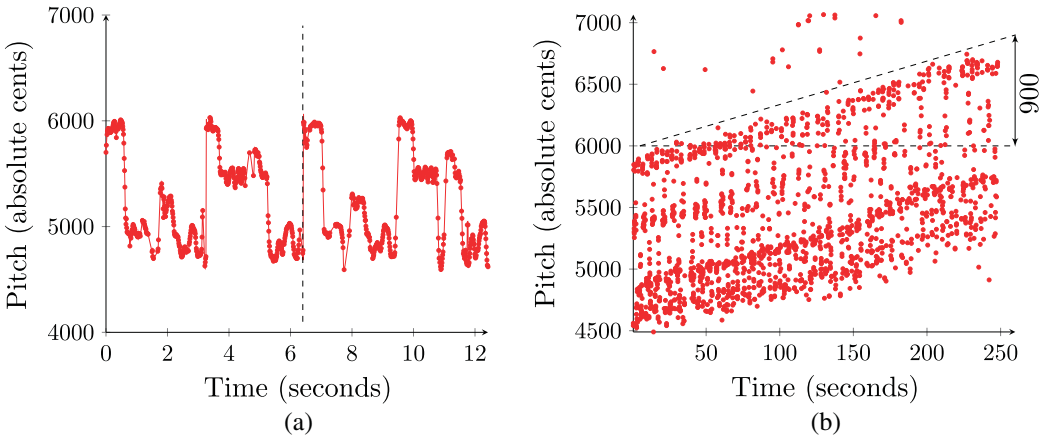


Fig. 12. A capella song performed by Nils Hotti from the Sami culture shows the gradual intentional pitch change during a song. The melodic motive however is constantly repeated (here shown twice).

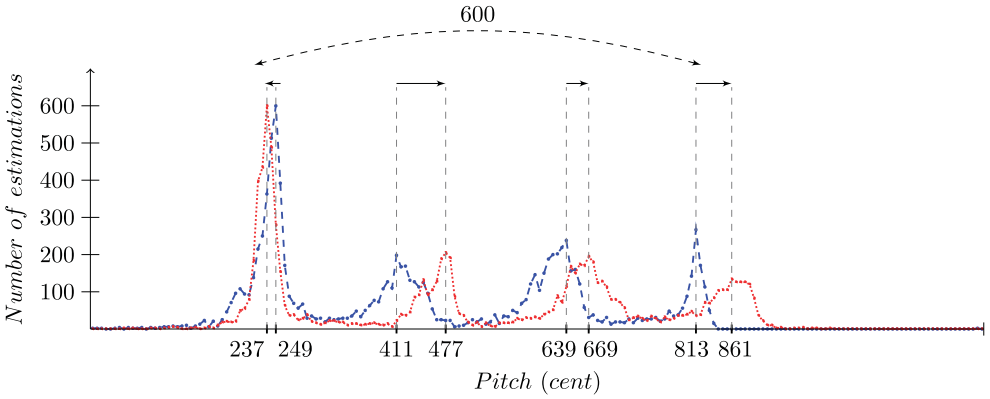


Fig. 13. Histogram of an African fiddle song. The second minute of the song is represented by the dashed line, the seventh minute is represented by the dotted line. The lowest, most stable pitch class is the result of the open string. It lost some tension during the piece and started to sound lower. This is in sharp contrast with the other pitch classes that sound higher, due to a change in hand position.

3.3 Shifted pitch distributions

Several difficulties in analysis and interpretation may arise due to pitch shift effects during musical performances. This is often the case with capella choirs. Figure 12 shows a nice example of an intentionally raised pitch, during solo singing in the Scandinavian Sami culture. The short and repeated melodic motive remains the same during the entire song, but the pitch rises gradually ending up 900 cents higher than the beginning. Retrieving a scale for the entire song is in this case irrelevant, although the scale is significant for the melodic motive. Figure 13 shows an example where scale organization

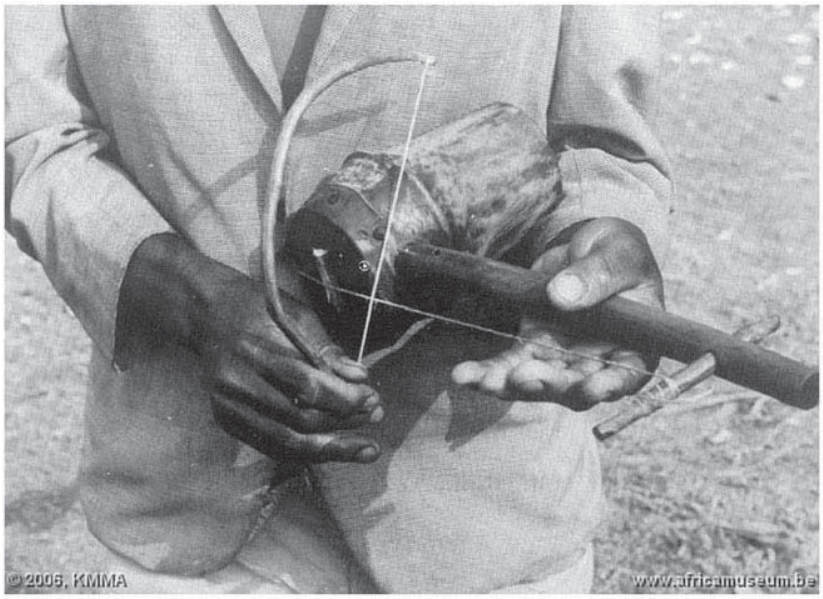


Fig. 14. The iningidi, a type of African fiddle. To play the instrument, the neck is held in the palm of the left hand so that the string can be stopped using the second phalanx of the index, middle and ring fingers. Consequently, a total of four notes can be produced.

depends on the characteristics of the instrument. This type of African fiddle, the iningidi, does not use a soundboard to shorten the strings. Instead the string is shortened by the fingers that are in a (floating) position above the string: an open string and three fingers give a tetratonic scale. Figure 14 shows an iningidi being played. This use case shows that pitch distributions for entire songs can be misleading, in both cases it is much more informative to compare the distribution from the first part of the song with the last part. Then it becomes clear how much pitch shifted and in what direction.

Interesting to remark on is that these intervals have more or less the same distance, a natural consequence of the distance of the fingers, and that, consequently, not the entire octave tessitura is used. In fact only 600 cents, half an octave, is used. A scale that occurs typically in fiddle recordings, which rather can be seen as a tetrachord. The open string (lowest note) is much more stable than the three other pitches that deviate more, as is shown by the broader peaks in the pitch class histogram. The hand position without soundboard is directly related to the variance of these three pitch classes. When comparing the second minute of the song with the seventh, one sees a clear shift in pitch, which can be explained by the fact the musician changed the hand position a little bit. In addition, another phenomena can be observed, namely, that while performing, the open string gradually loses tension, causing a small pitch lowering which can be noticed when comparing the two fragments. This is not uncommon for ethnic music instruments.

3.4 Tarsos' scripting applied to makam recognition

In order to make the use of scripting more concrete, an example is shown here. It concerns the analysis of Turkish classical music. In an article by Gedik and Bozkurt (2010), pitch histograms were used for—amongst other tasks—makam[8] recognition. The task was to identify which of the nine makams is used in a specific song. With the Tarsos API, a simplified, generalized implementation of this task was scripted in the Scala programming language. The task is defined as follows:

> For a small set of tone scales T and a large set of musical performances S, each brought in one of the scales, identify the tone scale t of each musical performance s automatically.

An example of makam recognition can be seen in Figure 15. A theoretical template—the dotted, red line—is compared to a pitch class histogram—the solid, blue line—by calculating the maximum overlap between the two. Each template is compared with the pitch class histogram, the template with maximum overlap is the guessed makam. Pseudocode for this procedure can be found in Algorithm 1.

Algorithm 1 Tone scale recognition algorithm.

```
1: T ← constructTemplates()
2: S ← fetchSongList()
3: for all s ∈ S do                          ▷ For all songs
4:    O ← {}                                 ▷ Initialize empty hash
5:    h ← constructPitchClassHisto(s)
6:    for all t ∈ T do                       ▷ For all templates
7:       o ← calculateOverlap(t, h)
8:       O[s] ← o                            ▷ Store overlap in hash
9:    end for
10:   i ← getFirstOrderedByOverlap(O))
11:   write s "is brought in tone scale" i
12: end for
```

[8] A makam defines rules for a composition or performance of classical Turkish music. It specifies melodic shapes and pitch intervals.

```
1  val makams = List( "hicaz","huseyni","huzzam","kurdili_hicazar",
                      "nihavend","rast","saba","segah","ussak")

   var theoreticKDEs = Map[java.lang.String,KernelDensityEstimate]()
   makams.foreach{ makam =>
6    val scalaFile = makam + ".scl"
     val scalaObject = new ScalaFile(scalaFile);
     val kde = HistogramFactory.createPichClassKDE(scalaObject,35)
     kde.normalize
     theoreticKDEs = theoreticKDEs + (makam -> kde)
11 }
```

Listing 1. Template construction.

```
   val directory = "/home/user/turkish_makams/"
   val audio_pattern = ".*.(mp3|wav|ogg|flac)"
   val audioFiles = FileUtils.glob(directory,audio_pattern,true).toList
4
   audioFiles.foreach{ file =>
     val audioFile = new AudioFile(file)
     val detectorYin = PitchDetectionMode.TARSOS_YIN.getPitchDetector(audioFile)
     val annotations = detectorYin.executePitchDetection()
9    val actualKDE = HistogramFactory.createPichClassKDE(annotations,15);
     actualKDE.normalize
     var resultList = List[Tuple2[java.lang.String,Double]]()
     for ((name, theoreticKDE) <- theoreticKDEs){
         val shift = actualKDE.shiftForOptimalCorrelation(theoreticKDE)
14       val currentCorrelation = actualKDE.correlation(theoreticKDE,shift)
         resultList = (name -> currentCorrelation) :: resultList
     }
     //order by correlation
     resultList = resultList.sortBy(_._2).reverse
19   Console.println(file + " is brought in tone scale " + resultList(0)._1)
   }
```

Listing 2. Makam recognition.

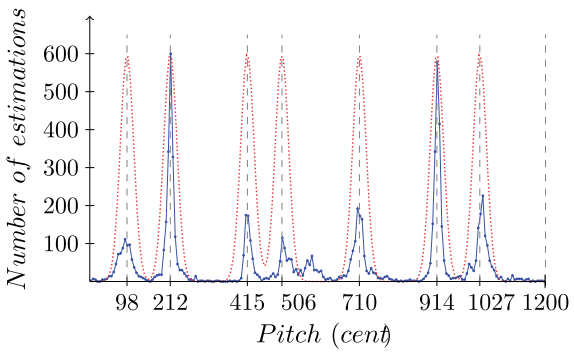


Fig. 15. The solid, blue line is a pitch class histogram of a Turkish song brought in the makam Hicaz. The dotted, red line represents a theoretical template of that same Hicaz makam. Maximizing the overlap between a theoretical and an actual pitch class histogram suggests which makam is used.

To construct the tone-scale templates theoretical descriptions of those tone scales are needed, for makams these can be found in Gedik and Bozkurt (2010). The pitch classes are converted to cent units and listed in Table 3. An implementation of $constructTemplates()$ in Algorithm 1 can be done as in Listing 1. The capability of Tarsos to create theoretical tone scale templates using Gaussian kernels is used, line 8. Line 7 shows how a scala file containing a tone-scale description is used to create an object with the same information, the height is normalized.

The $calculateOverlap(t, h)$ method from line 7 in Algorithm 1 is pitch invariant: it shifts the template to achieve maximum overlap with respect to the pitch class histogram. Listing 2 contains an implementation of the matching step. First a list of audio files is created by recursively iterating a directory and matching each file to a regular expression. Next, starting from line 4, each audio file is processed. The internal implementation of the YIN pitch detection algorithm is used on the audio file and a pitch class histogram is created (line 6,7). On line 10, normalization of the histogram is activated, to make the correlation calculation meaningful. In line 11 to line 15 the created histogram from the audio file is compared with the templates calculated beforehand (in Listing 1). The results are stored, ordered and eventually printed on line 19.

The script ran on Bozkurts data set with Turkish music: with this straightforward algorithm it is possible to correctly identify 39% of makams in a data set of about 800 songs. The results for individual makam recognition vary between 76% and 12% depending on how distinguishable the makam is. If the first three guesses are evaluated, the correct makam is present in 75% of the cases. Obviously, there is room for improvement by using more domain knowledge. A large improvement can be made by taking into account the duration of each pitch class in each template. Bozkurt does this by constructing templates by using the audio itself. A detailed failure analysis falls outside the scope of this article. It suffices to say that practical tasks can be scripted successfully using the Tarsos API.

4. Musicological aspects of Tarsos

Tarsos is a tool for the analysis of pitch distributions. For that aim, Tarsos incorporates several pitch extraction modules, has pitch distribution filters, audio feedback tools, and scripting tools for batch processing of large databases of musical audio. However, pitch distributions can be considered from different perspectives, such as ethnographical studies of scales (Schneider, 2001), theoretical studies in scale analysis (Sethares, 2005), harmonic and tonal analysis (Krumhansl & Shepard, 1979; Krumhansl, 1990), and other structural analy-

Table 3. The nine makams used in the recognition task.

Makam	**Pitch classes (in cents)**																					
Hicaz		113					384			498	701	792					996					
Huseyni			181		294					498	701				883		996					
Huzzam		113				316			430		701		812									1109
Kurdili Hicazar	90				294					498	701	792					996					
Nihavend				203	294					498	701	792					996					
Rast				203			384			498	701					905		1086				
Saba			181		294			407			701	792					996					
Segah		113				316				498	701			815								1109
Ussak			181		294					498	701	792					996					

Table 4. Results of the makam recognition task, using theoretical intervals, on the Bozkurt data set with Turkish music.

Makam	**Number of songs**	**Correct guesses**	**Percentage of correct guesses**
Kurdili Hicazar	91	32	35.16%
Huseyni	64	8	12.50%
Nihavend	75	27	36.00%
Segah	111	43	38.74%
Saba	81	50	61.73%
Huzzam	62	15	24.19%
Rast	118	23	19.49%
Ussak	102	54	52.94%
Hicaz	68	52	76.47%
Total	772	304	39.69%

sis approaches to music (such as set theoretical and Schenkerian). Clearly, Tarsos does not offer a solution to all these different approaches to pitch distributions. In fact, seen from the viewpoint of Western music analysis, Tarsos is a rather limited tool as it doesn't offer harmonic analysis, nor tonal analysis, nor even statistical analysis of pitch distributions. All of this should be applied together with Tarsos, when needed. Instead, what Tarsos provides is an intermediate level between pitch extraction (done by pitch extractor tools) and music theory. The major contribution of Tarsos is that it offers an easy to use tool for pitch distribution analysis that applies to all kinds of music, including Western and non-Western. The major contribution of Tarsos, so to speak, is that it offers pitch distribution analysis without imposing a music theory. In what follows, we explain why such tools are needed and why they are useful.

4.1 Tarsos and Western music theoretical concepts

Up to recently, musical pitch is often considered from the viewpoint of a traditional music theory, which assumes that pitch is stable (e.g. vibrato is an ornament of a stable pitch), that pitch can be segmented into tones, that pitches are based on octave equivalence, and that octaves are divided into 12 equal-sized intervals of each 100 cents, and so on. These assumptions have the advantage that music can be reduced to symbolic representations, a written notation, or notes, whose structures can be studied at an abstract level. As such, music theory has conceptualized pitch distributions as chords, keys, modes, sets, using a symbolic notation.

So far so good, but tools based on these concepts may not work for many nuances of Western music, and especially not for non-Western music. In Western music, tuning systems have a long history. Proof of this can be found in tunings of historical organs, and in tuning systems that have been explored by composers in the twentieth century (cf. Alois Haba, Harry Partch, Ivo Darreg, and Lamonte Young). Especially in non-Western classical music, pitch distributions are used that radically differ from the Western theoretical concepts, both in terms of tuning, as well as in pitch occurrence, and in timbre. For example, the use of small intervals in Arab music contributes to nuances in melodic expression. To better understand how small pitch intervals contribute to the organization of this music, we need tools that do not assume octave divisions in 12 equal-sized intervals (see Gedik & Bozkurt, 2010). Other types of music do not have octave equivalence (cf. the Indonesian gamalan), and also some music work with modulated pitch. For example, Henbing and Leman (2007) describe classical Chinese guqin music which uses tones that contain sliding patterns (pitch modulations), which form a substantial component of the tone and consider it as a succession of prototypical gestures. Krishnaswamy (2004a,b) introduces a set of 2D melodic units, melodic atoms, in describing Carnatic (South-Indian classical) music. They represent or synthesize

the melodic phrase and are not bound by a scale type. Hence, tools based on Western common music theoretical conceptions of pitch organization may not work for this type of music.

Oral musical traditions (also called ethnic music) provide a special case since there is no written music theory underlying the pitch organization. An oral culture depends on societal coherence, interpersonal influence and individual musicality, and this has implications on how pitch gets organized. Although oral traditions often rely on a peculiar pitch organization, often using a unique system of micro-tuned intervals, it is also the case that instruments may lack a fixed tuning, or that tunings may strongly differ from one instrument to the other, or one region to the other. Apparently, the myriad of ways in which people succeed in making sense out of different types of pitch organization can be considered as cultural heritage that necessitates a proper way of documentation and study (Moelants et al., 2009; Cornelis, Lesaffre, Moelants, & Leman, 2010).

Several studies have attempted developing a proper approach to pitch distributions. Gómez and Bonada (2008) look for pitch gestures in European folk music as an additional aspect to pitch detection. Moving from tone to scale research, Chordia, Rae (2007) acknowledge interval differences in Indian classical music, but reduce to a chromatic scale for similarity analysis and classification techniques. Sundberg and Tjernlund (1969) developed an automated method for extracting pitch information from monophonic audio for assembling the scale of the spilåpipa by frequency histograms. Bozkurt (2008), and Gedik and Bozkurt (2010) build a system to classify and recognize Turkish maqams from audio files using overall frequency histograms to characterize the maqams' scales and to detect the tonic centre. Maqams contain intervals of different sizes, often not compatible with the chromatic scale, but partly relying on smaller intervals. Moelants et al. (2007) focuses on pitch distributions of especially African music that deals with a large diversity of irregular tuning systems. They avoid a priori pitch categories by using a quasi-continuous rather than a discrete interval representation. In Moelants et al. (2009) they show that African songs have shifted more and more towards Western well temperament from the 1950s to the 1980s.

To sum up, the study of pitch organization needs tools that go beyond elementary concepts of the Western music theoretical canon (such as octave equivalence, stability of tones, equal temporal scale, and so on). This is evident from the nuances of pitch organization in Western music, in non-Western classical music, as well as in oral music cultures. Several attempts have been undertaken, but we believe that a proper way of achieving this is by means of a tool that combines audio-based pitch extraction with a generalized approach to pitch distribution analysis. Such a tool should be able to automatically extract pitch from musical audio in a culture-independent manner, and it should offer an approach to the study of pitch distributions and its relationship with tunings and scales. The envisioned tool should be able to perform this kind of analysis in an automated way, but it should be flexible enough to allow a musicologically grounded manual fine-tuning using filters that define the scope at which we look at distributions. The latter is indeed needed in view of the large variability of pitch organization in music all over the world. Tarsos is an attempt at supplying such a tool. On the one hand, Tarsos tries to avoid music theoretical concepts that could contaminate music that doesn't subscribe to the constraints of the Western music theoretical canon. On the other hand, the use of Tarsos is likely to be too limited, as pitch distributions may further draw upon melodic units that may require an approach to segmentation (similar to the way segmented pitch relates to notes in Western music) and further gestural analysis (see the references to the studies mentioned above).

4.2 Tarsos pitfalls

The case studies from Section 3 illustrate some of the capabilities of Tarsos as a tool for the analysis of pitch distributions. As shown Tarsos offers a graphical interface that allows a flexible way to analyse pitch, similar to other editors that focus on sound analysis (Sonic Visualizer, Audacity, Praat). Tarsos offers support for different pitch extractors, real-time analysis (see Section 2.4), and has numerous output capabilities (see Section 2.5). The scripting facility allows us to use Tarsos' building blocks in unique ways efficiently.

However, Tarsos-based pitch analysis should be handled with care. The following three recommendations may be taken into account: first of all, one cannot extract scales without *considering the music itself* . Pitch classes that are not frequently used, won't show up clearly in a histogram and hence might be missed. Also not all music uses distinct pitch classes: the Chinese and Indian music traditions have been mentioned in this case. Because of the physical characteristics of the human voice, voices can glide between tones of a scale, which makes an accurate measurement of pitch not straightforward. It is recommended to zoom in on the estimations in the melograph representation for a correct understanding.

Secondly, analysis of *polyphonic recordings should be handled with care* since current pitch detection algorithms are primarily geared towards monophonic signals. Analysis of homophonic singing for example may give incomplete results. It is advisable to try out different pitch extractors on the same signal to see if the results are trustworthy.

Finally, Schneider (2001) recognizes the use of 'pitch categories' but warns that, especially for complex inharmonic sounds, *a scale is more than a one dimensional series of pitches* and that spectral components need to be taken into account to get better insights in tuning and scales. Indeed, in recent years, it became clear that the timbre of tones and the musical scales in which these tones are used, are somehow related (Sethares, 2005). The spectral content of pitch (i.e. the timbre) determines the perception of consonant and dissonant pitch intervals, and therefore also the pitch scale, as the latter is a reflection of the preferred melodic and harmonic combinations of pitch. Based on the principle of minimal dissonance in pitch intervals, it is possible to derive pitch scales from

spectral properties of the sounds and principles of auditory interference (or critical bands). Schwartz and Purves (2004) argue that perception is based on the disambiguation of action-relevant cues, and they manage to show that the harmonic musical scale can be derived from the way speech sounds relate to the resonant properties of the vocal tract. Therefore, the annotated scale as a result of the precise use of Tarsos, does not imply the assignment of any characteristic of the music itself. It is up to the user to correctly interpret a possible scale, a tonal centre, or a melodic development.

4.3 Tarsos—future work

The present version of Tarsos is a first step towards a tool for pitch distribution analysis. A number of extensions are possible.

For example, given the tight connection between timbre and scale, it would be nice to select a representative tone from the music and transpose it to the entire scale, using a phase vocoder. This sound sample and its transpositions could then be used as a sound font for the MIDI synthesizer. This would give the scale a more natural feel compared to the general MIDI device instruments that are currently present.

Another possible feature is tonic detection. Some types of music have a well-defined tonic, e.g. in Turkish classical music. It would make sense to use this tonic as a reference pitch class. Pitch histograms and pitch class histograms would then not use the reference frequency defined in Appendix B but a better suited, automatically detected reference: the tonic. It would make the intervals and scale more intelligible.

Tools for comparing two or more scales may be added. For example, by creating pitch class histograms for a sliding window and comparing those with each other, it should be possible to automatically detect modulations. Using this technique, it should also be possible to detect pitch drift in choral, or other music.

Another research area is to extract features on a large data set and use the pitch class histogram or interval data as a basis for pattern recognition and cluster analysis. With a time-stamped and geo-tagged musical archive, it could be possible to detect geographical or chronological clusters of similar tone scale use.

On the longer term, we plan to add representations of other musical parameters to Tarsos as well, such as rhythmic and instrumental information, temporal and timbral features. Our ultimate goal is to develop an objective albeit partial view on music by combining those three parameters within an easy to use interface.

5. Conclusion

In this paper, we have presented Tarsos, a modular software platform to extract and analyse pitch distributions in music. The concept and main features of Tarsos have been explained and some concrete examples have been given of its usage. Tarsos is a tool in full development. Its main power is related to its interactive features which, in the hands of a skilled music researcher, can become a tool for exploring pitch distributions in Western as well as non-Western music.

References

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, *18*(9), 509–517.

Bozkurt, B. (2008). An automatic pitch analysis method for Turkish maqam music. *Journal of New Music Research (JNMR)*, *37*(1), 1–13.

Brossier, P. (2006). *Automatic annotation of musical audio for interactive applications* (PhD thesis). Queen Mary University of London, UK.

Cannam, C. (2008). *The Vamp Audio Analysis Plugin API: A Programmer's Guide.* Retrieved from: http://vamp-plugins.org/guide.pdf

Cannam, C., Landone, C., & Sandler, M. (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. Presented at *Open Source Software Competition, ACM,* Firenze, Italy.

Cha, S.-h. , (2007). Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, *1*(4), 300–307.

Chordia, P., & Rae, A. (2007). Raag recognition using pitch-class and pitch-class dyad distributions. In *Proceedings of the 8th International Symposium on Music Information Retrieval (ISMIR 2007),* Vienna, Austria, pp. 00–00.

Clarisse, L. P., Martens, J. P., Lesaffre, M., Baets, B. D., Meyer, H. D., & Leman, M. (2002). An auditory model based transcriber of singing sequences. In *Proceedings of the 3th International Symposium on Music Information Retrieval (ISMIR 2002),* Paris, France, pp. 116–123.

Cornelis, O., Lesaffre, M., Moelants, D., & Leman, M. (2010). Access to ethnic music: Advances and perspectives in content-based music information retrieval. *Signal Processing*, *90*(4), 1008–1031.

de Cheveigné, A., & Hideki, K. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, *111*(4), 1917–1930.

Gedik, A. C., & Bozkurt, B. (2010). Pitch-frequency histogram-based music information retrieval for Turkish music. *Signal Processing*, *90*(4), 1049–1063.

Gómez, E., & Bonada, J. (2008). Automatic melodic transcription of flamenco singing. In *Proceedings of 4th Conference on Interdisciplinary Musicology (CIM 2008),* Thessaloniki, Greece, pp. 00–00.

Henbing, L., & Leman, M. (2007). A gesture-based typology of sliding-tones in guqin music. *Journal of New Music Research (JNMR)*, *36*(2), 61–82.

Honingh, A., & Bod, R. (2011). In search of universal properties of musical scales. *Journal of New Music Research (JNMR)*, *40*(1), 81–89.

Klapuri, A. (2003). Multiple fundamental frequency estimation based on harmonicity and spectral smoothness. *IEEE Transactions on Speech and Audio Processing*, *11*(6), 804–816.
Krishnaswamy, A. (2004a). Melodic atoms for transcribing Carnatic music. In *Proceedings of the 5th International Symposium on Music Information Retrieval (ISMIR 2004),* Barcelona, Spain, pp. 345–348.
Krishnaswamy, A. (2004b). Multi-dimensional musical atoms in south-Indian classical music. In *Proceedings of the 8th International Conference on Music Perception & Cognition (ICMPC 2004),* Evanston, IL, pp. 00–00.
Krumhansl, C. L. (1990). Tonal hierarchies and rare intervals in music cognition. *Music Perception*, *7*(3), 309–324.
Krumhansl, C. L., & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 579–594.
Mearns, L., S. D., Benetos, E., & Dixon, S. (2011). Automatically detecting key modulations in J.S. Bach chorale recordings. In *Proceedings of the Sound Music and Computing Conference (SMC 2011),* Padova, Italy, pp. 00–00.
McLeod, P., & Wyvill, G. (2005). A smarter way to find pitch. In *Proceedings of the International Computer Music Conference (ICMC 2005),* Barcelona, Spain, pp. 00–00.
Moelants, D., Cornelis, O., & Leman, M. (2009). Exploring African tone scales. In *Proceedings of the 10th International Symposium on Music Information Retrieval (ISMIR 2009),* Kobe, Japan, pp. 00–00.
Moelants, D., Cornelis, O., Leman, M., Gansemans, J., Matthé, T., de Caluwe, R., de Tré, G., Matthé, T., & Hallez, A. (2007). Problems and opportunities of applying data- and audio-mining techniques to ethnic music. *Journal of Intangible Heritage*, *2*, 57–69.
Nobutaka, O., Miyamoto, K., Kameoka, H., Le Roux, J., Uchiyama, Y., Tsunoo, E., Nishimoto, T., & Sagayama, S. (2010). Harmonic and percussive sound separation and its application to MIR-related tasks. In Z.W. Ras and A. Wieczorkowska (Eds), *Advances in music information retrieval* (Studies in computational intelligence, Vol. 274). Berlin: Springer.
Schneider, A. (2001). Sound, pitch, and scale: From "tone measurements" to sonological analysis in ethnomusicology. *Ethnomusicology*, *45*(3), 489–519.
Schwartz, D. A., & Purves, D. (2004). Pitch is determined by naturally occurring periodic sounds. *Hearing Research*, *194*(1), 31–46.
Sethares, W. (2005). *Tuning timbre spectrum scale.* (2nd ed.). Berlin: Springer.
Six, J., & Cornelis, O. (2012). A robust audio fingerprinter based on pitch class histograms - applications for ethnic music archives. In *Proceedings of the Folk Music Analysis Conference (FMA 2012),* Seville, Spain.
Sundberg, J., & Tjernlund, P. (1969). Computer measurements of the tone scale in performed music by means of frequency histograms. *STL-QPS*, *10*(2–3), 33–35.
Tzanetakis, G., Ermolinskyi, A., & Cook, P. (2002). Pitch histograms in audio and symbolic music information retrieval. In *Proceedings of the 3th International Symposium on Music Information Retrieval (ISMIR 2002),* Paris, France, pp. 31–38.
von Helmholtz, H., & Ellis, A. J. (1912). *On the sensations of tone as a physiological basis for the theory of music* (translated and expanded by Alexander J. Ellis, 2nd. (English ed.). London: Longmans Green.

Appendix A. Pitch representation

Since different representations of pitch are used by Tarsos and other pitch extractors this section contains definitions of and remarks on different pitch and pitch interval representations.

For humans the perceptual distance between 220 and 440 Hz is the same as between 440 and 880 Hz. A pitch representation that takes this logarithmic relation into account is more practical for some purposes. Luckily there are a few:

MIDI *Note Number*

The MIDI standard defines note numbers from 0 to 127, inclusive. Normally only integers are used but any frequency f in Hz can be represented with a fractional note number n using Equation (A1).

$$n = 69 + 12\log_2\left(\frac{f}{440}\right), \tag{A1}$$

$$n = 12 \times \log_2\left(\frac{f}{r}\right); \quad r = \frac{440}{2^{(69/12)}} = 8.176\,\text{Hz}. \tag{A2}$$

Rewriting Equation (A1) to (A2) shows that MIDI note number 0 corresponds with a reference frequency of 8.176 Hz which is C_{-1} on a keyboard with A_4 tuned to 440 Hz. It also shows that the MIDI standard divides the octave into 12 equal parts.

To convert a MIDI note number n to a frequency f in Hz one of the following equations can be used.

$$f = 440 \times 2^{(n-69)/12}, \tag{A3}$$

$$f = r \times 2^{(n/12)} \text{ with } r = 8.176\,\text{Hz}. \tag{A4}$$

Using pitch represented as fractional MIDI note numbers makes sense when working with MIDI instruments and MIDI data. Although the MIDI note numbering scheme seems oriented towards Western pitch organization (12 semitones) it is conceptually equal to the cent unit which is more widely used in ethnomusicology.

Cent

von Helmholtz and Ellis (1912) introduced the nowadays widely accepted cent unit. To convert a frequency f in Hz to a cent value c relative to a reference frequency r also in Hz:

$$c = 1200 \times \log_2\left(\frac{f}{r}\right). \tag{A5}$$

With the same reference frequency r Equations (A5) and (A2) differ only by a constant factor of exactly 100. In an environment with pitch representations in MIDI note numbers and cent values it is practical to use the standardized reference frequency of 8.176 Hz.

To convert a frequency f in Hz to a cent value c relative to a reference frequency r also in Hz:

$$f = r \times 2^{(c/1200)}. \tag{A6}$$

Savart & Millioctaves
Divide the octave in 301.5 and 1000 parts respectively, which is the only difference with cents.

A.1 Pitch ratio representation

Pitch ratios are essentially pitch intervals, an interval of one octave, 1200 cents equal to a frequency ratio of $2/1$. To convert a ratio t to a value in cent c:

$$c = \frac{1200 \ln(t)}{\ln(2)}. \tag{A7}$$

The natural logarithm, the logarithm base e with e being Euler's number, is noted as ln. To convert a value in cent c to a ratio t:

$$t = e^{\frac{c \ln(2)}{1200}}. \tag{A8}$$

Further discussion on cents as pitch ratios can be be found in appendix B of Sethares (2005). There it is noted that:

> There are two reasons to prefer cents to ratios: Where cents are added, ratios are multiplied; and it is always obvious which of two intervals is larger when both are expressed in cents. For instance, an interval of a just fifth, followed by a just third is $(3/2)(5/4) = 15/8$, a just seventh. In cents, this is $702 + 386 = 1088$. Is this larger or smaller than the Pythagorean seventh $243/128$? Knowing that the latter is 1110 cents makes the comparison obvious.

A.2 Conclusion

The cent unit is mostly used for pitch interval representation while the MIDI key and Hz units are used mainly to represent absolute pitch. The main difference between cent and fractional MIDI note numbers is the standardized reference frequency. In our software platform Tarsos we use the exact same standardized reference frequency of 8.176 Hz which enables us to use cents to represent absolute pitch and it makes conversion to MIDI note numbers trivial. Tarsos also uses cents to represent pitch intervals and ratios.

Appendix B. Audio material

Several audio files were used in this paper to demonstrate how Tarsos works and to clarify musical concepts. In this appendix you can find pointers to these audio files.

The 30 s excerpt of the musical example used throughout Section 2 can be downloaded from http://tarsos.0110.be/tag/

JNMR and is courtesy of: WERGO/Schott Music & Media, Mainz, Germany, www.wergo.de and Museum Collection Berlin. Ladrang Kandamanyura (slendro pathet manyura) is track eight on Lestari—*The Hood Collection, Early Field Recordings from Java*—SM 1712 2. It was recorded in 1957 and 1958 in Java.

The yoiking singer of Figure 12 can be found on a production released on the label Caprice Records in the series of Musica Sveciae Folk Music in Sweden. The album is called Jojk CAP 21544 CD 3, Track No 38 Nila, hans svager/His brother-in-law Nila.

The API example (Section 3.4) was executed on the data set by Bozkurt. This data set was also used in Gedik and Bozkurt (2010). The Turkish song brought in the makam Hicaz from Figure 15 is also one of the songs in the data set.

For the comparison of different pitch trackers on the pitch class histogram level (Section 3.2) a subset of the music collection of the Royal Museum for Central Africa (RMCA, Tervuren, Belgium) was used. We are grateful to the RMCA for providing access to its unique archive of Central African music. A song from the RMCA collection was also used in Section 3.1. It has the tape number MR.1954.1.18-4 and was recorded in 1954 by missionary Scohy-Stroobants in Burundi. The song is performed by a singing soloist, Léonard Ndengabaganizi. Finally the song with tape number MR.1973.9.41-4, also from the collection of the RMCA, was used to show pitch shift within a song (Figure 13). It is called *Kana nakunze* and was recorded by Jos Gansemans in Mwendo, Rwanda in the year 1973.

Cornelis, O., Six, J., Holzapfel, A., & Leman, M. (2013). Evaluation and Recommendation of Pulse and Tempo Annotation in Ethnic Music. Journal of New Music Research. 42 (2) pp. 131-149.

Journal of New Music Research, 2013
Vol. 42, No. 2, 131–149, http://dx.doi.org/10.1080/09298215.2013.812123

Routledge
Taylor & Francis Group

Evaluation and Recommendation of Pulse and Tempo Annotation in Ethnic Music

Olmo Cornelis[1,*] , Joren Six[1] , Andre Holzapfel[2] and Marc Leman[3]

[1]*University College Ghent, Belgium;* [2]*Bahçeşehir University, Turkey;* [3]*Ghent University, Belgium*

Abstract

Large digital archives of ethnic music require automatic tools to provide musical content descriptions. While various automatic approaches are available, they are to a wide extent developed for Western popular music. This paper aims to analyse how automated tempo estimation approaches perform in the context of Central-African music. To this end we collect human beat annotations for a set of musical fragments, and compare them with automatic beat tracking sequences. We first analyse the tempo estimations derived from annotations and beat tracking results. Then we examine an approach, based on mutual agreement between automatic and human annotations, to automate such analysis, which can serve to detect musical fragments with high tempo ambiguity.

1. Introduction

In an effort to preserve the musical heritage of various cultures, large audio archives with ethnic music have been created at several places throughout the world.[1] With the widespread availability of digital audio technology, many archiving institutions have started to digitize their audio collections to facilitate better preservation and access.[2] Meanwhile, a good number of audio collections have been fully digitized, which enables the next step to make these audio archives more accessible for researchers and general audiences.

Computational Ethnomusicology from this perspective, aims at providing better access to ethnic audio music collections using modern approaches of content-based search and retrieval (Tzanetakis, Kapur, Schloss & Wright, 2007; Cornelis, Lesaffre, Moelants & Leman, 2010). This research field has its roots in Western Musicology, as well as in Ethnomusicology and Music Information Retrieval. Current computational tools for the content-based analysis of Western musical audio signals are well established and have begun to reach a fair performance level as seen in many applications, publications and the MIREX initiative. However, for the field of ethnic music, it is still unclear which computational tools for content-based analysis can be applied successfully. Given the diversity and oral character of ethnic music, Computational Ethnomusicology faces many challenges. A major difficulty is concerned with the influence and dominance of Western musical concepts in content-based analysis tools. It is generally believed that the influence of Western concepts may affect the interpretation of the extracted audio features. However, there is little information about the exact nature of this possible contamination. It may be that tools based on low-level acoustical features perform reasonably well, while tools that focus on higher-level musical concepts perform less well. In this context, one could question whether existing beat tracking and tempo extraction tools, typically developed and tested on, mainly, Western music, can be readily applied to African music.

In this paper, we focus on tools for beat tracking and tempo extraction from Central-African music. The overall[3] aim of this study is to see to what extent meaningful results can be expected from the automatic tempo analysis of Central-African music. The research in this paper relies on existing computational tools, and does not aim to introduce novel approaches in beat tracking and tempo estimation. A useful byproduct of this research could be a new way to identify

[1]British Library (London), CREM and SDM (Paris), Ethnologisches Museum (Berlin), RMCA (Brussels), Essen Folksong Collection (Warsaw), GTF (Vienna), and many more.

[2]See Appendix C for a number references to digitization projects.

[3]The Music Information Retrieval Evaluation eXchange (MIREX) is an annual evaluation campaign for Music Information Retrieval (MIR) algorithms. More info about MIREX can be found on http://www.music-ir.org.

Correspondence: Olmo Cornelis, University College Ghent, School of Arts, Hoogpoort 64, 9000 Ghent, Belgium.
E-mail: olmo.cornelis@hogent.be

ethnic music with ambiguous tempo relations and reveal information of a higher metrical hierarchy: from beats to meter.

Our goal is to explore whether a set of 17 automatic beat trackers and tempo estimators (i) can be used as a tool for extracting tempo from Central-African musical audio, (ii) can give insight into the ambiguity of tempo perception, (iii) can detect problematic cases for tempo annotation, and (iv) if it can provide information about a higher metrical level.

In order to be able to evaluate the performance of the beat trackers, we compare them with the performance of 25 professional musicians, who manually annotated the beat for 70 audio fragments. The results of both human and computational annotations are analysed and compared with each other. The goal is to see how large the variability is in both sets of annotations (automatic and manual) and whether ambiguity in human annotations implies ambiguity in computational annotations, and how well the two match.

The paper is structured as follows; Section 2 presents aspects of tempo in music. Section 3 gives an overview of related literature. Section 4 outlines our methodology and describes the used data collection. Section 5 contains the results of these experiments. Section 6 elaborates on considerations in the field of approaching ethnic music. Section 7 concludes the paper.

2. On the concept of tempo

Willenze (1964) points out the relationship between the measurable, or objective time and the time that is experienced, the subjective time. This reflects the traditional distinction between the theoretical tempo that is implied in a score, and the tempo that comes out of performance. Although the score written by a composer is handled as a primary source, musical notation in the case of transcription is typically considered to be a subjective assessment of the transcriber. Especially in the area of ethnic music this has been mentioned several times, as for example in the work of Brandel (1961).

Subjective assessments of tempo in music are determined by studying synchronization with the pulse. However, at least in Western music, the pulse often functions within a larger structure that is called the meter. Lerdahl and Jackendoff (1983) speak about strong and weak beats (*instances of a pulse*) and they approach meter as a super structure on top of *'a relatively local phenomenon'*. The perception of pulse and meter is associated with a perceivable regularity that creates expectations in a time span. For this reason, one can tap along with any music that has a regular/repetitive basis. Therefore, meter facilitates the structuring of the beats over time.

Non-Western rhythmical phenomena are different from Western rhythmical phenomena. Ethnomusicologists tend to recognize the concept of pulse that organizes music in time, but they assess the structuring of pulses in a way that is different from the concept of meter. From all their theories and concepts, the idea of the *fastest pulse* as a basis for understanding aspects of timing seems to be the most fundamental, general, and useful, as it allows the widest variety of interpretations. In this context, Arom (1985) states that African music is not based on bars, which define the meter as in classical music, but on pulsations, a succession of isochronous time units.

Thus, rather than using the concept of meter, the structuring of pulses is based on the concept of sequences, forming the starting point for further analysis of rhythms. The best-known approach is the Time Unit Box System (TUBS) notation, developed by Kubik (1994) and Koetting (1970) for annotating West African drums. It is a graphical annotation approach that consists of boxes of equal length put in horizontal sequence. Each box represents an instance of the fastest pulse in a particular musical piece. If an event occurs, the box is marked, if not the box is left empty. TUBS are most useful for showing relationships between layers of complex rhythms. An example of this notation can be found in Figure 3, Section 5.1.

The approach of rhythmical organization by Kubik (1994) and Koetting (1970) is based on three levels. The first level is the elementary pulsation, a framework of fast beats that define the smallest regular units of a performance as an unheard grid in the mind of the performer. The second level is formed by a subjective reference beat. There are no preconceived strong or weak parts of the meter, and the beats are often organized in a repetitive grid of 3, 4, 6, 8 or 12 units. The point of departure is so ingrained that it needs no special emphasis. For this reason, the first beat is often acoustically veiled or unsounded. For outsiders this can cause a phase shift. On top of these two levels, Kubik adds a third level, which he calls the cycle. A cycle would typically contain 16 to 48 beats. The introduction of numbered cycles (Kubik, 1960) replaced conventional Western time signatures in many transcriptions of African music. The main advantage of conceiving these large cycles is that polymeter structures resolve in it.

Agawu (2003) introduced *topoi*, which are short distinct, memorable rhythmic figures of modest duration that serve as a point of temporal reference. The presence of these repetitive topoi shows that there is an underlying pulse. He writes that '*West and Central African dances feature a prominently articulated, recurring rhythmic pattern that serves as an identifying signature*'. Seifert, Schneider and Olk (1995) followed a similar path of the smallest pulse as basis for a theoretical and integrated research strategy for the interpretation of non-Western rhythmical phenomena, based on the TUBS of Kubik and Koetting.

Connected to the idea of the fastest pulse, Jones (1959) was the first to describe the *asymmetric structure* of the higher rhythmical patterns. A well-known common example of such a pattern is the 12-beat pattern that contains a seven and a five stroke component, of which one is prevalent while its complementary pattern is latent, and is tapped as a syncopated pulse. The pattern appears later as an example in Section 5 and is illustrated there by Figure 3.

Another prominent rhythmical phenomenon in African music are *interlocking patterns*. They consist of two or more (rhythmic or melodic) lines that have different starting points, running one smallest beat apart from each other. Kubik suggests that the origin of these interlocking patterns could have

initiated from pestle-pounding strokes by two or three women that alternately strike in a mortar. The patterns are fundamental to much African music.

A final remark concerns a call by Agawu (1995) for rebalancing the presumed importance of rhythmical elements in African music over the other musical parameters. Agawu (2003) believes that the rhythmical elements and their organization in African music are over-conceptualized. In his writings he lists, quotes, and reviews many of the great ethnomusicologists' ideas of the 20th century. Contrary to these ideas, he suggests a more explorative bottom-up approach and he warns ethnomusicologists against the eagerness of constructing African music as essentially different from the West.

This shows that the concepts of pulse, meter, and tempo are still a topic of discussion, and that this discussion should be taken into account when trying to apply computational content-based analysis methods to Central-African music.

3. Literature on tapping experiments

Apart from concepts on pulse, meter, sequences, and tempo, it is also of interest to consider experiments on tapping. Experiments on synchronized finger tapping along with beat of the music (Desain & Windsor, 2000; Large, 2000; Wohlschlager & Koch, 2000; Moelants & McKinney, 2004; Repp, 2006) reveal some interesting aspects that should be taken into account when studying beat and tempo in Central-African music.

One aspect concerns the range in which musical tempo can be perceived, namely, between 200 to 1500 ms (milliseconds), or 40 to 300 Beats Per Minute (bpm) (Pöppel, 1978; Moelants & McKinney, 2004). In cases of slower tempi one tends to subdivide, while faster tempi physically cannot be performed. Within that space, Moelants mentions there is a *preferred tempo-octave* lying between 81 and 162 bpm.

It is perhaps superfluous to mention that the regularity of beats is never strictly rigid. In musical performances as well as in human synchronization tapping tasks, minor deviations are present in the signal and data, but these are inherent to musical and to human performance. They do not influence the global tempo, but are characteristics of the microtiming in the music. A related aspect concerns the *negative asynchrony* (Repp, 2006), the phenomenon that subjects tend to tap earlier than the stimulus (typically between 20 and 60 ms), which shows that subjects perform motor planning, and thus rely on anticipation, during the synchronization task (Dixon, 2002).

Another aspect concerns *tempo octaves*, the phenomenon that subjects tend to synchronize their taps with divisions or multiplications of the main tempo. These tempo octaves are regularly reported and they are the main argument to identify a tempo as being *ambiguous*. Indeed, the human perceivable tempo limitations (40–300 bpm) span a large range of tempi, namely, more or less three tempo-octaves. Consequentially, the listener has different possibilities in synchronizing (tapping) with the music. Therefore, ambiguity arises in the tempo annotations of a group of people. These choices are related to personal preference, details in the performance, and temporary mood of the listener (Moelants, 2001). This subjectivity has large consequences in approaching tempo and meter in a scientific study. McKinney and Moelants (2006) demonstrate that for pieces with tempi around 120 bpm, a large majority of listeners are very likely to perceive this very tempo, whereas faster and slower tempi induce more ambiguity, with responses spread over two tempo-octaves (Moelants & McKinney, 2004). This connects to the *2 Hz resonance theory of tempo perception* (Van Noorden & Moelants, 1999), according to which tempo perception and production is closely related to natural movement, with humans functioning as a resonating system with a natural frequency. The preferred tempo is located somewhere between 110 and 130 bpm, and therefore creates a region in which music is tapped less ambiguously (Moelants, 2002).

In this perspective, it is possible to distinguish between beat rate and/or tapping rate on the one hand, and the perceived tempo on the other hand (Epstein, 1995). The beat rate is the periodicity which best affords some form of bodily synchronization with the rhythmic stimulus. It may or it may not directly correspond to the perceived tempo, especially when the latter is considered as a number that reflects a rather complex *Gestalt* that comes out of the sum of musical factors, combining the overall sense of a work's themes, rhythms, articulations, breathing, motion, harmonic progressions, tonal movement, and contrapuntal activity. As such, the beat could be different from the perceived tempo. Early research by Bolton (1894) reported already the *phenomenal grouping* as an aspect of synchronized tapping; when he presented perfectly isochronous and identical stimuli to subjects they spontaneously subdivided, by accentuation, into units of two, three, or four. London (2011) speaks of *hierarchically-nested periodicities* that a rhythmic pattern embodies. The observation of subdivisions and periodicity brings Parncutt (1994) to the question what *phase* listeners tend to synchronize to when listening to music and what cues in the musical structure influence these decisions.

Another aspect concerns the *ambiguity of meter perception* (McKinney & Moelants, 2006). In music theory, the meter of a piece is considered as an unambiguous factor, but some music could be interpreted both with a binary or a ternary metric structure. Handel and Oshinsky (1981) presented a set of polyrhythmic pulses and asked people to synchronize along with them. The general outcome was that 80% of the subjects tapped in synchrony with one of the two pulses, whereas 12% of the subjects tapped the co-occurrence of the two pulses, and 6% tapped every second or third beat. The choice of preferred pulse however was not clear. A conclusion was that subjects tend to follow the fastest of the two pulses that make the polyrhythm when the global tempo is slow, and that subjects tend to follow the slowest pulse in a fast global tempo. When the global tempo is too high, people switch to a lower tempo octave. If the presented polyrhythm consists of different pitch content, the lower pitch element was the preferred frequency. Finally, Handel and Oshinsky concluded that if the tempo of

the presented series of beats is very high, the elements are temporally so tightly packed that the pulse becomes part of the musical foreground instead of the pulsation that is part of the musical background. For polyrhythms, this transition point is about 200 ms or 300 bpm.

The above overview shows that research on synchronized tapping tasks has to take into account several aspects that are likely to be highly relevant in the context of Central-African stimuli where we typically deal with complex polyrhythms.

4. Methodology

4.1 Experiment 1: Human

4.1.1 Procedure: Tap along

Tempo annotation is the ascription of a general tempo to a musical piece, expressed in beats per minute (bpm). Beat synchronization is the underlying task for the identification of a basic pulse from which the tempo is derived. Subjects were asked to *tap to the most salient beat* of the audio fragments. More information on the stimuli can be found in Section 4.1.2. For each tap annotation containing taps at the time instances $t_1, ..., t_N$ (s), we obtain a set of $N - 1$ inter-tap distances $D = d_1, ..., d_{N-1}$ (s). Then, a tempo in bpm is assigned to the piece by calculating the median of $60/D$.

The experiment was done on a laptop with the subjects listening to the audio fragments on headphones while tapping on the keyboard space bar. Since manual annotation of tempo is an intense and time consuming task, the data was recorded in two sessions with a small pause between the two. Subjects could restart any fragment if they had doubts about their annotation. The number of retries and the tapping data for each retry were recorded together with the final tapping data. All the data was organized and recorded by the software Pure Data. To ensure that the data is gathered correctly a test with a click track was done, with the interval between the clicks being constantly 500 ms. The average tapping interval was 499.36 ms, with a standard deviation of 20 ms. The low standard deviation implies that the measurement system has sufficient granularity for a tapping experiment.

4.1.2 Stimuli: Audio fragments

The stimuli used in the experiment were 70 sound fragments, each with a length of 20 s, selected from the digitized sound archive of the Royal Museum for Central Africa (RCMA), Tervuren, Belgium. The archive of the Department of Ethnomusicology contains at present about 8000 musical instrument and 50,000 sound recordings, with a total of 3000 h of music, most of which are field recordings made in Central Africa with the oldest recordings dating back to 1910. The archive has been digitized not only to preserve the music but also to make it more accessible (Cornelis et al., 2005). Results of the digitization project can be found at http://music.africamuseum.be. The 70 fragments were chosen from the RMCA archive. It was attempted to cover a wide range of tempi and to include only fragments without tempo changes. The songs contained singing, percussion, and other musical instruments, in soloist or in group performances. This set of 70 stimuli will be referred to as *fragments* in the subsequent sections.

4.1.3 Participants: Musicians

The experiment was carried out by 25 participants. All of them were music students at the University College Ghent – School of Arts (Belgium), who were expected to play, practice, and perform music for several hours a day. The group consisted of 14 men and 11 women, ranging in age from 20 to 34 years.

4.2 Experiment 2: Software

Within the Music Information Retrieval community automated tempo estimation and beat tracking are important research topics. While the goal of the former is usually the estimation of a tempo value in bpm, the latter aims at estimating a sequence of time values that coincides with the beat of the music. Beat tracking and tempo estimation are applied in diverse applications, such as score alignment, structure analysis, play-list generation, and cover song identification. This paper however does not compare or evaluate such algorithmic approaches. For these matters, please refer to Gouyon et al. (2006), Zapata and Gómez (2011), and the yearly MIREX competition[4].

Automatic tempo analysis was done on the stimuli by a set of 17 beat trackers and tempo estimation algorithms (see Appendix B). All parameters for each algorithm were left on the default values and no adaption to the stimuli was pursued. Some algorithms only give an ordered list of tempo suggestions (Beatcounter, Mixmeister, Auftakt), here only the primary tempo annotation was considered. For the beat tracking algorithms, a tempo estimation was derived from the beat sequences in the same way as for the human taps as described in Section 4.1. To be able to compare the results of the automatic tempo analysis with the human annotations, the same stimuli were used as in the first experiment (see Section 4.1.2).

4.3 Comparison: Measuring beat sequence/annotation agreement

Recently, a method based on mutual agreement measurements of beat sequences was proposed by Holzapfel, Davies, Zapata, Lobato Oliveira and Gouyon (2012). This method was applied for the automatic selection of informative examples for beat tracking evaluation. It was shown that the Mean Mutual Agreement (MMA) between beat sequences can serve as a good indicator for the difficulty of a musical fragment for either automatic or human beat annotation. A threshold on MMA could be established above which beat tracking was assumed to be feasible to a subjectively satisfying level. For

[4]http://www.music-ir.org

the beat sequence evaluation in this paper, five out of the 17 algorithms were selected (Klapuri, Eronen & Astola, 2006; Dixon, 2007; Ellis, 2007; Oliveira, Gouyon, Martin & Reis, 2010; Degara et al., 2011). This selection was made for several reasons. First, some of the 17 approaches are pure tempo estimators that give only tempo values in bpm, and not beat sequences. Second, in Holzapfel et al. (2012) it was shown that this selection increases diversity and accuracy of the included beat sequences, and, third, this selection guarantees comparability with results presented in Holzapfel et al. (2012).

Comparing beat sequences is not a straightforward task; two sequences should be considered to agree not only in the case of a perfect fit, but also in the presence of deviations that result in perceptually equal acceptable beat annotations. Such deviations include small timing deviations, tempi related by a factor of two, and a phase inversion (off-beat) between two sequences, to name only the most important factors that should not be considered as complete disagreement. Because of the difficulty of assessing agreement between beat sequences, various measures have been proposed that differ widely regarding their characteristics (Davies, Degara & Plumbley, 2009). In this paper we restrict ourselves to two evaluation measures that are suitable for the two tasks at hand, which are spotting complete disagreement between sequences and investigating the types of deviations between sequences.

(1) Information Gain (Davies, Degara & Plumbley, 2011): Local timing deviations between beat sequences are summarized in a beat error histogram. The beat error histogram is characterized by a concentration of magnitudes in one or a few bins if sequences are strongly related, and by a flatter shape if the two sequences are unrelated. The deviation of this histogram from the uniform distribution, the so-called 'information gain', is measured using K-L divergence. The range of values for Information Gain is from 0 to 5.3 bits, with the default parameters as proposed in Davies et al. (2011). This measure punishes completely unrelated sequences with a value of 0 bits, while all sequences with some meaningful relation tend to score higher. Such meaningful relations include a constant beat-relative phase shift, or simple integer relations between the tempi of the sequences. This means that off-beat or octave differences do not lead to a strong decrease in this measure. The maximum score can only be reached when all beat errors between the two sequences fall into the same beat error histogram bin, with the bin width being, for example 12.5 ms at 120 bpm. MMA measured with this measure will be denoted as MMA_D.

(2) F-measure: A beat in one sequence is considered to agree with the second sequence if it falls within a ± 70 ms tolerance window around a beat in the second sequence. Let the two sequences have $|A|$ and $|B|$ beats, respectively. We denote the number of beats in the first sequence that fall into such a window of the second sequence as $|A_{win}|$, and the number of beats in the second sequence that have a beat of the first sequence in their tolerance window as $|B_{win}|$. Note that if several beats of the first sequence fall into one tolerance window, $|A_{win}|$ is only incremented by one. Then the F-measure is calculated as

$$F = \frac{2 * P * R}{P + R} \tag{1}$$

with $P = |A_{win}|/|A|$ and $R = |B_{win}|/|B|$. The F-measure has a range from 0% to 100% and drops to about 66% when two sequences are related by a factor of two, while a value of 0% is usually only observed when two sequences have the exact same period, but a phase offset. Note that two unrelated sequences do not score zero but about 25% (Davies et al., 2009). MMA measured with this measure will be denoted as MMA_F.

We will investigate, how many fragments in the RMCA subset can be successfully processed with automatic beat tracking, and to what extent the human annotations correlate with the estimated beat sequences. For this task MMA_D will be applied, as it was shown in Holzapfel et al. (2012) to reliably spot difficult musical fragments. For the fragments, which were judged to be processable by automatic beat tracking, we will apply MMA_F, as we can differentiate which types of errors occurred for a given fragment. For example, values of 66% are mostly related to octave relations between the compared sequences, and an off-beat relation is in practice the only case which results in a value of 0%.

The MMA values for a fragment will be obtained by computing the mean of the $N(N-1)/2$ mutual agreements, with $N = 5$ for beat trackers, and $N = 25$ for human annotations. We will differentiate between beat sequences, which are obtained from algorithms (referred to as BT), and tapped annotations from human annotators (referred to as TAP).

5. Results

5.1 Human tempo annotations

In Appendix A we list the tempo annotations for all songs and all annotators. We assigned a general tempo value to each song by choosing the tempo that most people tapped. A tempo was considered similar if it did not deviate by more than 6 bpm from the assigned tempo. The other tempi were considered in relation to this assigned tempo, and could be divided into tempo octaves (half, double, triple tempo), related tempi (usually a mathematical relation with the assigned tempi), related octaves (half, double, triple of the related tempo), unrelated tempi (no relation with the assigned tempo). Also some people tapped annotations of different length creating a pattern as, for example, $2 + 3$ in a meter of 5 and $2 + 3 + 3$ for some songs

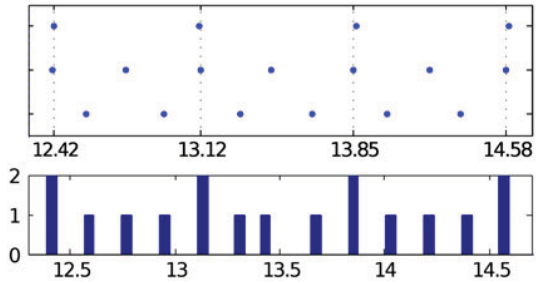


Fig. 1. Small fragment (Track 25) of tapped onsets of three persons, one following the tempo octave (tempo halving), and two persons in different phase. Histogram below.

in 8, and those were specified as patterns without attempting to derive a tempo value from them.

A first glance at the results, Table 1, shows that 68 songs could be assigned a general tempo, two songs had such wide range of tempi that no general tempo could be assigned. They were both capella vocal songs, which contained rather recitation than singing. Of the remaining 68 songs, only two songs were labelled unanimously. For 64 songs people tapped tempo octaves, and also for 43 songs related tempi were present. For the songs that had both octaves and related tempi, the distribution was equal: 19 songs had more octaves than related tempi, and 19 songs had more related tempi than octaves. This last group, which formed 27%, can be seen as songs with high ambiguity in tempo perception. These songs contained several instruments that combined polymetric layers. People tended to have distributed preference in following different instruments.

Table 2 lists the distribution of all 1750 annotations: 60% correspond to the assigned tempo, 17% correspond to tempo octaves, while only 9% correspond to related tempi. Apparently, in many songs (61%) some people do hear related tempi, but mostly this is a small group of people. But, even after applying a threshold on the minimum number of relation occurrances as in Table 3, 23% of the songs were still tapped in a related tempo by five or more persons (from the 25). This shows that related tempi are not coincidental or individual cases, but that a quarter of the audio set had tempo ambiguity, similar to what was derived in the previous paragraph.

The individual differences on the median over the 70 songs was remarkable, with personal medians ranging from 77 up to 133 bpm. In affirmation with some elements from the literature, there is indeed a large agreement on tempo annotations in the region 120–130 bpm, namely 83% (10% tapped a tempo octave, and only 2% tapped a related tempo for this tempo region). Eight of the 10 songs in this tempo region were tapped with a binary meter. In the other tempi regions, ambiguity was much higher, but the set was too small to deduce tendencies. What was noticed is that songs around 90 bpm received only a few tempo octaves, but more related tempi.

When we focus on the properties of individual songs, the pieces with a meter in five deserve special attention. The annotations were very diverse, and can be divided into different groups. Some people tapped exactly on the fastest pulse, while others only tapped each fifth beat of this pulse, creating a tempo range of five tempo octaves. Some people tapped every second beat of the fastest pulse level, which implies going 'on and off beat' per bar, creating an alternating syncopation. Several people tapped a subdivided pattern of 2 and 3 beats and some people tapped every 2.5 beats, subdividing the meter of five into two equal parts. This diversity reoccurred for each song that had a meter in five.

Agawu mentions that cultural insiders easily identify the pulse. For those who are unfamiliar with such specific culture, and especially if the dance or choreographic movements cannot be observed, it can be difficult to locate the main beat and express it in movement (Agawu, 2003). De Hen (1967) considers that rhythm is an alternation of tension and relaxation. The difference between Western music and African music, he writes, lies in the opposite way of counting, where Western music counts heavy-light and African music is the other way around. The human annotations support these points. Figure 1 zooms in on a tap annotation where persons 1 and 2 tap the same tempo but in a different phase. Figure 2 visualizes a sim-

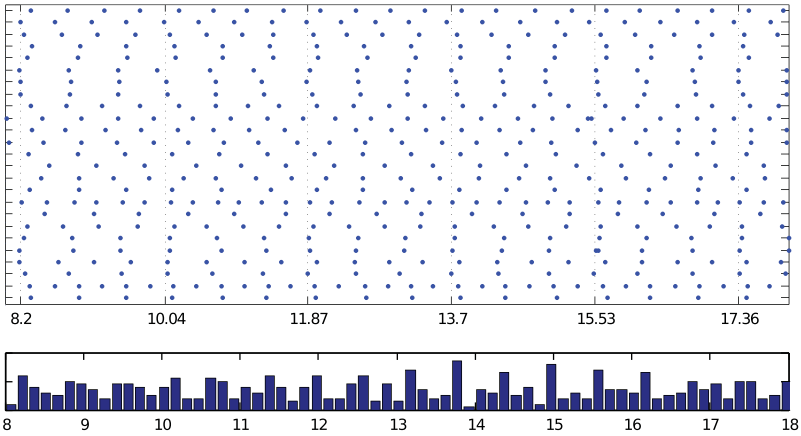


Fig. 2. Fragment of track 61, where the group is divided in binary and ternary tapping. Two people follow the smallest pulse (tempo doubling). Time indications were manually added to mark the bars. The histogram shows this polymetric presence.

Table 1. Overview of audio fragments organized by sorts of human assigned tempi.

Type	#	%	Track IDs
Unanimous tempo	2	2.9%	5, 56
+ Tempo Octaves (no related)	23	32.9%	4, 6, 7, 8, 9, 10, 13, 14, 15, 17, 23, 25, 35, 42, 44, 50, 51, 55, 57, 58, 60, 65, 70
Tempo octaves < Related tempi	19	27.1%	28, 1, 62, 22, 20, 59, 63, 18, 41, 66, 53, 54, 37, 43, 52, 26, 39, 19, 64
Tempo octaves = Related tempi	3	4.3%	29, 34, 45
Tempo octaves > Related tempi	19	27.1%	69, 32, 38, 48, 61, 30, 33, 40, 24, 27, 47, 68, 12, 31, 67, 36, 49, 11, 3
+ Related tempi (no octaves)	2	2.9%	2, 46
No tempo	2	2.9%	16, 21
Total number of records	70		

Table 2. Distribution of all annotations (1750 human annotations, 1190 BT tempi) over available classes.

Type	Human (%)	BT (%)
Identical	60%	48%
Octave	17%	18%
Related	9%	19%
Related Tempo Octave	3%	3%
Unrelated	9%	6%
Pattern	2%	0%

Table 3. Distribution of all annotations over available classes if a threshold is set.

		At least one		More than One		More than Two		More than Five	
Human	Tempo octaves	64	91%	56	80%	44	63%	28	40%
	Related tempo	43	61%	32	46%	25	36%	16	23%
	Related octave	25	36%	13	19%	7	10%	1	1%
	Pattern	37	53%	24	34%	15	21%	10	14%
	Unrelated tempo	19	27%	11	16%	6	9%	2	3%
BT	Identical	64	94%	61	90%	58	85%		
	Octave	52	76%	41	60%	28	41%		
	Related	52	76%	38	56%	28	41%		
	Related Octave	18	26%	9	13%	5	7%		
	Unrelated tempo	31	46%	20	29%	13	19%		

Table 4. BT annotations organized by meter and their classification along the human tempo references.

Meter	Identical	Octave	Related
(1)	1 25 58 60	0	0
2	5 20 27 31 34 35 43 44 47 51 53	42 64 70	3
3	2 10 12 26 33 40 45	0	19 29 36 38 59 61
4	4 6 9 13 15 17 18 23 37 50 54 55 56 69	8 14 30 57	24 67
5	22 41 46 49	52	11
6	7 28 32 39 48 66	63	62 65 68

ilar example where the binary annotations vary in phase. This specific fragment was very ambiguous—13 persons tapped ternary, 10 binary—what is especially remarkable is that the group of ternary people synchronize in phase, while the binary annotations differ much more. It is clear that the ambiguity is not only between binary and ternary relations, but that there is a phase ambiguity as well. As an explorative case study, a small group was asked to write down the rhythmical percussive ostinato pattern from an audio fragment. The result shown in Figure 3 is striking by its variance. At first sight it seems so incomparable one would even question if they were listening to the same song. To summarize, it appears

that people perceive different tempi, different meter, different starting points and assign different accents and durations to the percussive events.

As a final insight, we have transposed the idea of the TUBS notations (Time Unit Box System) to the human annotations (see Section 2). While TUBS is most useful for showing relationships between complex rhythms, it is here used for visualizing the annotation behaviour where the place of the marker in the box indicates the exact timing of the tapped event. Hence, it visualizes the human listeners' synchronization to music. In Figure 4, a fragment of the tapped annotations is given. One sees clearly that there is quite some variance in trying to synchronize with the music, although the global tempo was unambiguous. This variance is mainly caused by the individual listeners tapping stably but in different phases than the others.

5.2 Tempo annotation by beat trackers

The tempo annotations of the 17 beat trackers (BTs) are listed in Appendix B; each column containing the tempo estimates of each song.

The reference tempo for evaluating the tempo estimations was the tempo that most people tapped (see Appendix A). As with the analysis of the human annotations, the other categories were: tempo octaves, related tempo, related tempo octaves and unrelated tempi. The category of patterns was left out, as beat tracking algorithms are designed to produce a regular pulse.

In most cases the majority of the 17 BTs did match the tempo assigned by humans, namely for 46 fragments (67.6%), listed in Table 4. For nine songs the tempo octave was preferred by the beat trackers, in most instances (seven), they suggested the double tempo. For the remaining 13 songs, the beat trackers preferred the related tempo above the assigned tempo, 10 times they preferred the binary pulse for the ternary pulse tapped by humans, and only two times the ternary for the binary. One instance concerned a meter of five where the tempo estimation of the BT split up the meter into 2.5. Looking at Table 3, the assigned tempo was detected by at least one BT in 64 songs (94%), and by three of the five BTs still in 58 songs (85%).

Table 2 contains the distribution of the 1190 annotations which are comparable to the overall human annotations. At 48%, there is a slight decrease in identical tempo annotations, while the category of the related tempi increases up to 19%.

We can conclude that the beat trackers give a reliable result: two thirds of the tempi were analysed identically to the human annotations. For the other songs the majority of the BTs suggested a tempo octave or a related tempo. In songs with higher ambiguity (where people assigned several tempi), it appears that the BTs tend to prefer binary meter over ternary, and higher tempi over slower. The preference for higher tempo is also reflected in the medians for each beat tracker over the 70 songs, with a range of 109–141 bpm, and one outlier

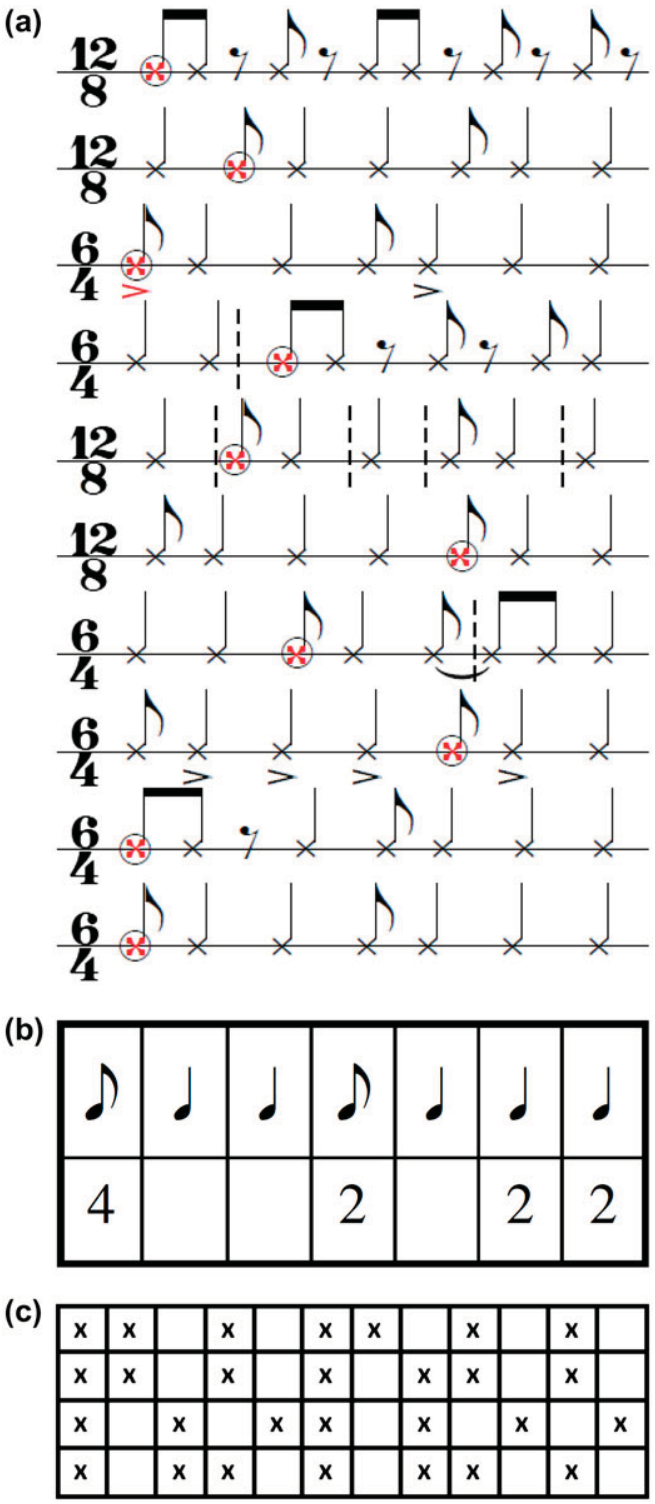


Fig. 3. Different transcriptions of the widespread asymmetrical 12-pulses ostinato rhythmical pattern/timeline. (a) Different transcriptions of the same rhythmical pattern derived from listening to a song (in case MR.1973.9.19-2A) by 10 people. The circled note indicates same place in the shifted pattern. (b) Number of transcriptions at different starting points in the pattern. (c) TUBS notation of the general pattern with four different starting points.

of 191 bpm, higher than the human medians mentioned in Section 5.1.

5.3 Human annotations versus beat trackers

As a first step we determined all mutual agreements between the five beat trackers that are contained in our committee, using the Information Gain measure (see Section 4.3). In Figure 5 the histograms of these mutual agreements for all musical fragments in the RMCA subset are shown, where the histograms are sorted by their MMA_D value. It can be observed that there is an almost linear transition from histograms with concentration at low agreement values to histograms with very high agreements on the right side of Figure 5. The vertical red line marks the threshold for perceptually satisfying beat sequences (MMA=1.5 bits), which was established in listening tests (Zapata, 2012). Out of the 70 fragments in the dataset 57 lie on the right side of this threshold, which implies that for 81% of this data at least one of the five beat sequences can be considered as perceptually acceptable. This percentage is higher than the one reported for a dataset of Western music (73%, Zapata (2012)). In the previous section we showed that 59 songs have either correct or half/double tempo. That proportion is quite close to the 81% we measure here.

We will show the difference between songs having beat sequences with low MMA and those having a high MMA between their sequences using two examples. One example was taken from the left side of the red line in Figure 5 and the other from the right side of it. An excerpt of the beat sequences for the low-MMA_D song is shown in Figure 6. It is apparent that the beat sequences are largely unrelated, both in terms of tempo as well as in terms of phase alignment. On the other hand, in Figure 7 the song with high MMA_D has beat sequences that are more strongly related. Their phase is well aligned, however, there are octave relationships between the tempi of the beat sequences. This can also be seen from the TUBS representation, which is less randomly distributed than for the low-MMA_D song depicted in Figure 6. This clarifies that by calculating the MMA_D we can obtain an estimation about the agreement between beat sequences or annotations without the necessity of a time-consuming manual analysis.

When directing our attention towards the human annotations, we obtain an unexpected result. In Figure 8 it can be seen that from low agreement among beat sequences follows low agreement among human annotations, which can be seen by the population of the lower-left rectangle formed by the 1.5 bit threshold lines. However, high agreement among beat trackers does not imply high agreement among human tappers; a significant amount of fragments with a BT-MMA_D above the threshold has quite low TAP-MMA_D values (lower-right rectangle). This is quite different from the result for Western music presented in Holzapfel et al. (2012), where this quadrant was not populated at all, indicating that good beat tracker performance always implied high agreement among human tappers. Inspection of the human annotations related to the fragments in the lower-right quadrant revealed that they are indeed characterized by a large variability for each fragment. The audio for these fragments appears to have several polyrhythmic layers, almost independent polyphony, often with flute, rattle, singing, and dense percussion. Several fragments in the lower quadrants contained rattles, which have an unclear attack, resulting in poorly aligned tapped sequences.

From the 12 fragments in the lower-left quadrant only one had a binary meter while six of them were ternary. Two were in five and three were undefined. From the 11 fragments in the lower-right quadrant, the meters were equally distributed, but for this selection the average tempo stands out with 140 bpm, whereas it was 102 bpm for the lower-left quadrant and 109 bpm for the upper quadrants. The BT tempi follow the same tendency, but less distinct. The upper quadrants had an average of 17 persons tapping the same tempo, while the lower quadrants 12. When we add the number of the retries of the human annotations, which can indicate the more difficult files since people were in doubt with their first annotation, we see a very large portion of the retries appearing in the lower-left quadrant. For the lower-right quadrant, some fragments barely had any retries while others had many. There was no relation between meter and retries, except for the meter in five which apparently needed one or more retries from most people.

As we can now determine for which fragments some meaningful relation can be found in a set of beat sequences or annotations by using MMA_D, we now go one step further and explore which kind of tempo relations might be encountered between these sequences, and if there are off-beat relationships. For example, in Figure 7 we saw a set of beat sequences that are well aligned in phase, but were characterized by octave relationships. To this end we will analyse the MMA_F, which results in characteristic values in the presence of specific tempo and phase relations, as explained in Section 4.3. For the 57 fragments above the MMA_D threshold in Figure 5 we depict the MA histograms obtained using the F-measure in Figure 9, sorted again by MMA_D. Hence, this plot represents the BT-mutual agreement histograms of the same fragments as on the right side of the red line in Figure 5, but the histograms are computed using the F-measure. The curve on the right side of the histograms depicts the sum of each bin over all 57 fragments. We can see that the largest amount of sequences agree perfectly (100%). The peak close to 66% is mainly caused by sequences that are well aligned but have tempo relations of factor two. High values in the histogram at zero help identifying sets of sequences with identical tempi, but phase shifted relations. Our example shown in Figure 7 which contained octave errors finds itself in column 42 of the image in Figure 9. It has a large peak in the bin related to 66% which can be seen by the black spot in that area. Hence, by observing the shape of a histogram (i.e. a single column in Figure 9), we can obtain valuable insight into what relations exist between an arbitrary set of beat sequences or annotations. While tempo relations between regular sequences can easily be obtained by determining the relations between their average inter beat distances, this says nothing about the accuracy of their align-

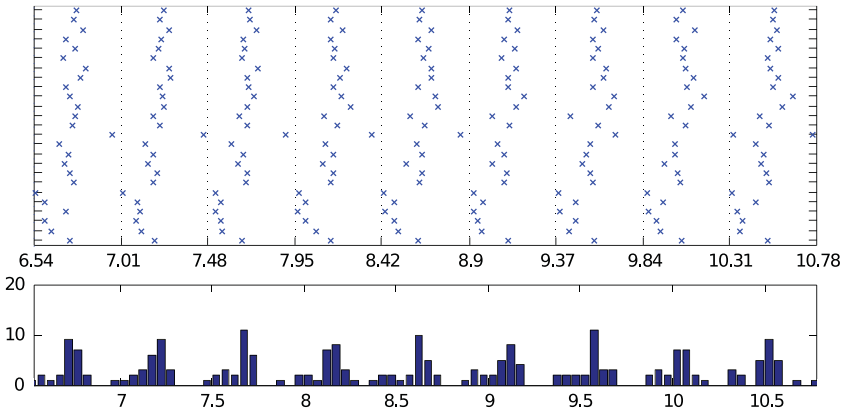


Fig. 4. Fragment of track 56 where each box represents one beat, as in a TUBS representation. The unanimously assigned tempo however conceals large time differences in human onsets. The dotted lines are manually added as a reference.

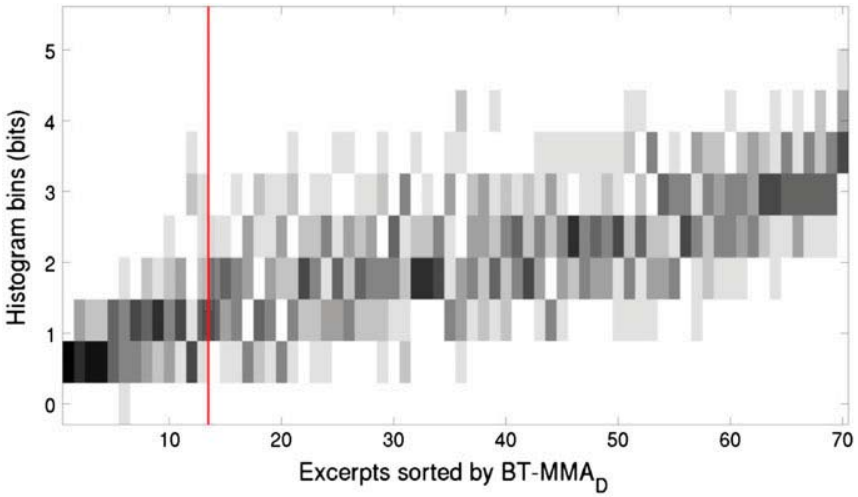


Fig. 5. Each column of the image depicts a histogram obtained from $5 * 4/2$ mutual agreements of the five beat sequences for each song in the RMCA subset. The histograms are sorted by their mean values (BT-MMA). Dark colours indicate high histogram values. The dotted red line marks the threshold above which a perceptually satisfying beat estimation can be performed.

ment in phase. Thus, examining the existence of peaks in the F-measure MA histograms can give a better understanding about this alignment. Furthermore, these histograms have the property that they give an even more accurate representation when the number of compared sequences is high. This is quite helpful, as for a large number of sequences manual analysis gets more and more difficult. While we showed example sequences for beat tracking algorithm outputs, such insight can also be obtained for human annotations.

6. MIR and ethnic music discussion

6.1 Awareness on possible biased approaches

Most music software applications, interfaces, and underlying databases are optimized for descriptions related to Western popular music. A common practise of such music information retrieval software is to take the musical characteristics and semantic descriptions of Western music as a standard, and to develop tools that are based upon a series of Western cultural concepts and assumptions. These assumptions apply to structural aspects (e.g. tonality, assumption of octave equivalence, instrumentation), social organization of the music (e.g. composers, performers, audience) and technical aspects (e.g. record company, release date). For non-Western music however, there is no guarantee that these concepts can be easily applied (Tzanetakis et al., 2007). On the contrary, imposing Western concepts onto non-Western music can lead to incorrect or incomplete information. The predominant focus on the composer and performer illustrates this typically Western approach, whereas in non-Western music this information is often unknown or even irrelevant. In turn, non-Western music often has a very specific function, such as working song, rowing, hunting, which is an unfamiliar concept for Western music. There is, however, a need for reorienting methodolo-

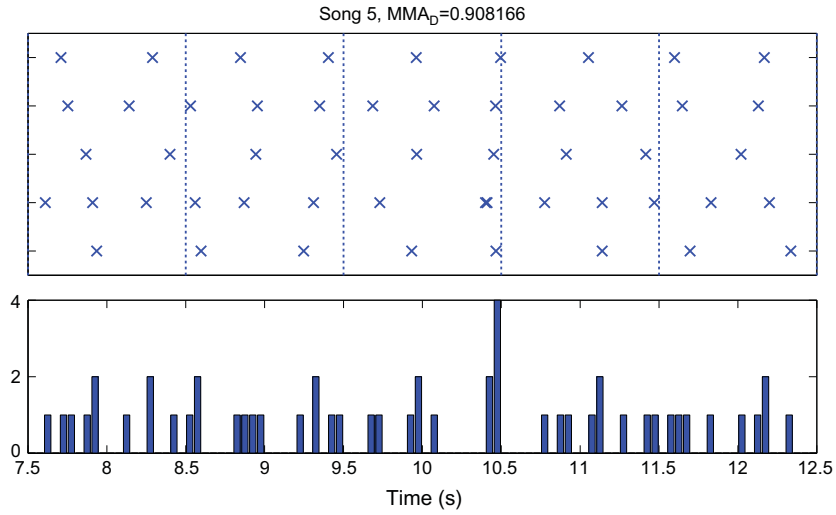


Fig. 6. Beat sequences of the five beat trackers in the committee for a song with low MMA_D.

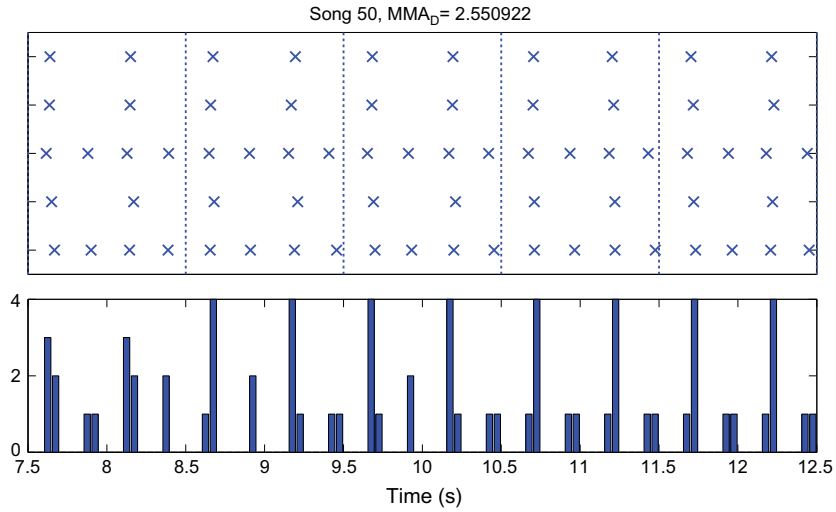


Fig. 7. Beat sequences of the five beat trackers in the committee for a song with high MMA_D.

gies since over the last decade several national and European projects were launched which aim at digitization of musical libraries, with at least a part of ethnic music. See Appendix C for a limited list of such as projects.

On the other hand, as can be seen in the results of this research, the beat tracking software does perform well, even without any specific fine-tuning towards the set of Central-African music. The paradigm of focusing on the smallest pulse, as some ethnomusicologists suggest, is an effective starting point which the beat trackers are capable of.

6.2 Transcription

A general concern is the indirect relation between the sounding music, its written representation, and the musical intentions of the composer/performer as described by Leman (2007). This relationship is even weaker in the context of ethnic music. Any musical performance is an intense and individual interpretation of its performers' knowledge and history. The ethnomusicologist who listens to this musicalized language faces an immense challenge if he wants to (re)produce scores starting from the audio as it is heard.

In such tasks, transcription has since long been the first step before studying an oral culture. Often a transcription relies on Western notation, sometimes specially invented symbols are added, and some others prefer to use graphical visualization of the audio. More about the complexities of transcriptions can be found in Nettl (1983), two chapters by Ter Ellingson in Myers (1993) and the chapter *Notation and Oral Tradition* by Shelemay (2008).

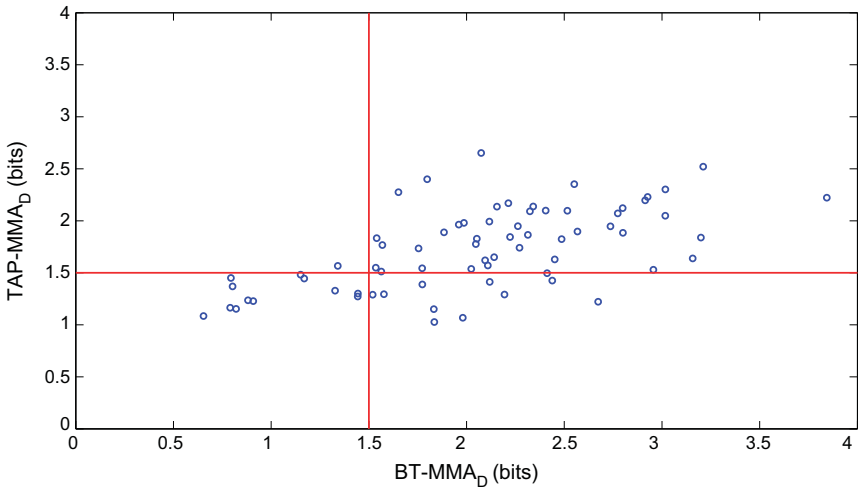


Fig. 8. Scatter plot of the MMA_D values obtained from human tappings and the beat tracking algorithms. Red lines indicate the 1.5 bit threshold.

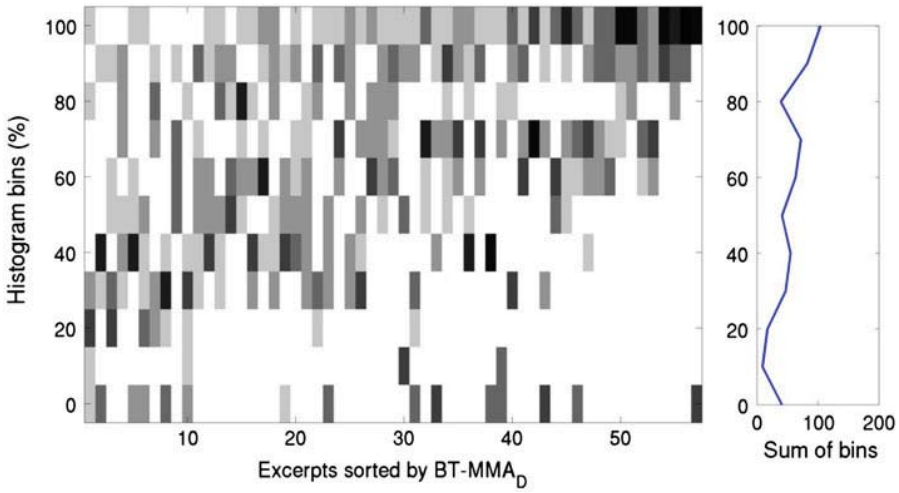


Fig. 9. Each column of the image depicts a histogram obtained from $5 * 4/2$ mutual agreements of the five beat sequences for each song in the RMCA audio subset, measured with the F-measure. The histograms are sorted by the BT-MMA$_D$. Dark colours indicate high histogram values. On the right we see the result of summing up each bin over all histograms.

As a final note, we identify some polarizing issues: namely the descriptive notation, a meticulously detailed notation that tries to capture every aspect of the audio but makes it hard to read or even understand, versus the prescriptive transcription that merely consists of the information needed by the insider (Nettl, 1983). And secondly, in the context of African music which is of very repetitive nature, one can ask if a full transcription is needed, or that it is allowed to summarize the song to its essential components by filtering out small variations (Wade, 2009).

With the aim of developing automated tools for transcription, one must be aware of all these elements. They set out rules that should not be seen as a additional difficulties, rather they should be seen as guidelines which a multidisciplinary approach of musicology, ethnomusicology and computer engineering should follow.

7. Conclusions and future work

This paper presents the preliminary research on the development of a computational approach for analysing temporal elements in ethnic music. For a good understanding of tempo in ethnic music, a case study with Central-African music was conducted. Both human annotations, and the output of a set of beat trackers were compared to discover insights in the tempo estimations results, in the computational potential, and in some perceptual phenomena themselves. Tempo is based on the regular and repetitive pulse of music, and will form a basis for any further analysis, annotation and transcription. The experiment showed the ambiguity in perception of tempo and meter, both for humans and for beat trackers. The beat trackers obtained comparable results with the human annotations, with a slight tendency to prefer binary pulsation in ambiguous

situations and to prefer a higher tempi octave. We also found a notable ambiguity in phase indication.

Gathering multiple beat trackers entails some advantages: if their results are combined, they appear to detect temporal ambiguity in songs where humans showed a similar perception. Detecting such information is important for the user, as it is, after all, our intention to create a realistic analysis platform where the user makes the final decision on any annotation or transcription. The software only makes suggestions that can be followed, adapted or ignored. Another interesting advantage is that the combination of the several tempo estimations does tell us something about the temporal organization behind the pulsation: combining the group of tempo estimations can give suggestions about the metrical organization of the piece.

The given hypotheses can be affirmed by this research: (i) a set of BTs can be used as a reliable method for tempo extraction in Central-African music with results comparable with human annotations, (ii) the set of BTs gives similar insights into the ambiguity of tempo perception as in human tempo perception, and (iii) the set of BTs does mostly detect problematic cases for tempo annotation. The fourth hypothesis seems promising namely that the combined results of the set of BTs can provide information of a higher metrical level, but this has not been investigated further in a computational way.

It is the intention to add the proposed approach into the existing software package Tarsos (Six & Cornelis, 2011), which currently is focused on analysis of pitch organization in ethnic music.

Acknowledgements

This research was supported by the University College Ghent and by the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

We are grateful to the RMCA (Royal Museum for Central Africa) in Belgium for providing access to its unique archive of Central African music.

Finally we would like to thank José R. Zapata for his support in running beat tracking algorithms.

References

Agawu, K. (1995). The invention of African rhythm. *Journal of the American Musicological Society*, *48*, 380–395.

Agawu, K. (2003). *Representing African Music*. New York: Routledge.

Arom, S. (1985). *African Polyphony and Polyrhythm: Musical Structure and Methodology*. Cambridge: Cambridge University Press.

Bolton, T. (1894). Rhythm. *American Journal of Psychology*, *6*, 145–238.

Brandel, R. (1961). *The Music of Central Africa*. Leiden: Martinus Nijhof.

Cornelis, O., De Caluwe, R., Detré, G., Hallez, A., Leman, M., & Matthé, T. (2005). Digitisation of the ethnomusicological sound archive of the RMCA. *IASA Journal*, *26*, 35–44.

Cornelis, O., Lesaffre, M., Moelants, D., & Leman, M. (2010). Access to ethnic music: Advances and perspectives in content-based music information retrieval. *Signal Processing*, *90*, 1008–1031.

Davies, M. E. P., Degara, N., & Plumbley, M. D. (2009). *Evaluation Methods for Musical Audio Beat Tracking Algorithms*, Technical Report No. C4DM-TR-09-06. Queen Mary University of London, Centre for Digital Music.

Davies, M. E. P., Degara, N., & Plumbley, M. D. (2011). Measuring the performance of beat tracking algorithms using a beat error histogram. *IEEE Signal Processing Letters*, *18*, 157–160.

Degara, N., Argones, E., Pena, A., Torres, M., Davies, M. E. P., & Plumbley, M. D. (2011). Reliability-informed beat tracking of musical signals. *IEEE Transactions on Audio, Speech and Language Processing*, *1*, 290–301.

De Hen, F. (1967). De muziek uit Afrika. *Bulletin d'information de la coopération au développement, 14*, 8–14.

Desain, P., & Windsor, L. (2000). *Rhythm, Perception and Production*. The Netherlands: Swets & Zeitlinger.

Dixon, S. (2002). Pinpointing the beat: Tapping to expressive performances. In *International Conference on Music Perception and Cognition (ICMPC 2002), Sydney, Australia, July 17–21* (CD ROM, 4pp.). Adelaide, Australia: Causal Productions.

Dixon, S. (2007). Evaluation of the audio beat tracking system BeatRoot. *Journal of New Music Research (JNMR)*, *36*, 39–50.

Ellis, D. P. W. (2007). Beat tracking by dynamic programming. *Journal of New Music Research (JNMR)*, *36*, 51–60.

Epstein, D. (1995). *Shaping Time: Music, the Brain, and Performance*. New York: Schirmer.

Gouyon, F., Klapuri, A., Dixon, S., Alonso, M., Tzanetakis, G., & Uhle, C. (2006). An experimental comparison of audio tempo induction algorithms. *IEEE Transactions on Speech and Audio Processing*, *14*(5), 1832–1844.

Handel, S., & Oshinsky, J. (1981). The meter of syncopated auditory polyrhythms. *Percept Psychophys*, *1*, 1–9.

Holzapfel, A., Davies, M. E. P., Zapata, J. R., & Gouyon, F. (2012). Selective sampling for beat tracking evaluation. *IEEE Transactions on Audio, Speech and Language Processing*, *20*, 2539–2548.

Jones, A. (1959). *Studies in African Music*. London: Oxford University Press.

Klapuri, A. P., Eronen, A. J., & Astola, J. T. (2006). Analysis of the meter of acoustic musical signals. *IEEE Transactions On Audio Speech And Language Processing*, *14*, 342–355.

Koetting, J. (1970). Analysis and notation of West African drum ensemble. *Selected Reports in Ethnomusicology*, *1*, 115–147.

Kubik, G. (1960). The structure of Kiganda xylophone music. *African Music*, *2*, 6–30.

Kubik, G. (1994). *Theory of African music*. Chicago: University of Chicago Press.

Large, E. W. (2000). On synchronizing movements to music. *Human Movement Science*, *19*, 527–566.

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press.

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

London, J. (2011). Tactus ≠ tempo: Some dissociations between attentional focus, motor behavior, and tempo judgment. *Empirical Musicology Review*, *6*, 43–55.

McKinney, M. F., & Moelants, D. (2006). Ambiguity in tempo perception: What draws listeners to different metrical levels? *Music Perception*, *24*, 155–166.

Moelants, D. (2001). *Een model voor ritmeperceptie toegepast op de muziek van de 20ste eeuw* (PhD thesis). University Ghent, Belgium.

Moelants, D. (2002). Preferred tempo reconsidered. *Proceedings of the 7th International Conference on Music Perception and Cognition (ICMPC 2002), Sydney, Australia, July 17–21* (pp. 580–583). Causal Productions: Adelaide, Australia.

Moelants, D., & McKinney, M. F. (2004). Tempo perception and musical content: What makes a piece fast, slow or temporally ambiguous? In: S. D. Lipscomb, R. Ashley, R. O. Gjerdingen & P. Webster (Eds.), *Proceedings of the International Conference on Music Perception and Cognition (ICMPC 2004), Evanston, IL* (pp. 558–562). Adelaide, Australia: Causal Productions.

Myers, H. (1993). *Ethnomusicology, historical and regional studies*. Abingdon: Routledge.

Nettl, B. (1983). *The Study of Ethnomusicology, 31 Issues and Concepts*. Champaign, IL: University of Illinois Press.

Oliveira, J., Gouyon, F., Martin, L., & Reis, L. (2010). IBT: A realtime tempo and beat tracking system. In *Proceedings of the 11th International Symposium on Music Information Retrieval (ISMIR 2010)*, Utrecht, Netherlands, pp. 291–296.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, *11*, 409–464.

Pöppel, E. (1978). Time perception. In R. Held, H. Teuber and H. Leibowits (Eds), *Handbook of Sensory Physiology, Vol. 8: Perception* (Ch. 23). Berlin: Springer Verlag.

Repp, B. (2006). Musical synchronization. In: E. Altenmuller, M. Wiesendanger & J. Kesselring (Eds.), *Music, Motor Control and the Brain* (pp. 55–76). Oxford: Oxford University Press.

Seifert, U., Schneider, A., & Olk, F. (1995). On rhythm perception: Theoretical issues, empirical findings. *Journal of New Music Research (JNMR)*, *24*, 164–195.

Shelemay, K. (2008). Notation and oral tradition. In: R. M. Stone (Ed.), *The Garland Handbook of African Music* (pp. 24–44). New York: Routledge.

Six, J., Cornelis, O., & Leman, M. (2013). Tarsos, a Modular Platform for precise Pitch Analysis of Western and non-Western music. *Journal of New Music Research (JNMR)*, doi: 10.1080/09298215.2013.797999

Tzanetakis, G., Kapur, A., Schloss, W. A., & Wright, M. (2007). Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*, *1*(2), 1–24.

Van Noorden, L., & Moelants, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research*, *28*, 43–66.

Wade, B. C. (2009). *Thinking Musically*. Oxford: Oxford University Press.

Willenze, T. (1964). *Algemene Muziekleer (Aula-boeken; 644)*. Dordrecht: Het Spectrum.

Wohlschlager, A., & Koch, R. (2000). Synchronisation error: An error in time perception. In: P. Desain & W. L. Windsor (Eds.), *Rhythm Perception and Production* (pp. 115–127). The Netherlands: Swets & Zeitlinger.

Zapata, J. R., & Gómez, E. (2011). Comparative evaluation and combination of audio tempo estimation approaches. In *Proceedings of the 42nd AES Conference on Semantic Audio, Ilmenau, Germany, July 22–24* (pp. 198–207). New York: Audio Engineering Society.

Zapata, J. R., Holzapfel, A., Davies, M. E. P., Lobato Oliveira, J., & Gouyon, F. (2012). Assigning a confidence threshold on automatic beat annotation in large datasets. In *Proceedings of the 13th International Symposium on Music Information Retrieval (ISMIR 2012)*, Porto, Portugal, pp. 157–162.

Appendix A: Human annotations

Table A1. Human tempo annotations on a set of seventy sound fragments.

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	T21	T22	T23	T24	T25	T26	T27	T28	T29	T30	T31	T32	T33	T34	T35
1	**112**	**83**	**105**	59	**119**	**129**	75	**83**	**94**	66	38	**114**	72	**81**	**120**	103	**101**	58	**65**	**111**	104	75	69	77	83	61	65	93	*38*	*95*	72	*72*	67	**98**	**128**
2	54	*27*	52	**117**	**122**	167	**150**	**81**	47	**133**	83	*38*	**144**	**80**	60	49	195	66	85	**109**	49	117	71	*51*	83	0	258	55	**114**	**63**	105	252	*92*	102	63
3	**119**	**81**	*79*	**118**	**122**	**129**	**150**	161	**95**	**133**	81	**117**	**146**	**79**	**120**	88	**101**	103	133	**109**	112	*117*	**140**	*103*	158	60	**128**	94	105	*185*	**144**	*140*	*90*	**98**	**126**
4	**112**	66	52	**117**	**120**	65	**150**	**80**	**94**	66	40	**115**	72	**80**	**120**	54	**100**	108	**67**	**110**	105	74	70	**152**	**162**	**122**	199	92	**115**	**62**	**145**	*140*	*45*	50	**126**
5	**114**	**81**	52	59	**122**	**128**	**152**	**81**	**95**	66	103	58	**146**	40	**120**	57	**100**	*49*	**65**	**109**	54	**150**	71	*51*	**167**	**120**	**129**	57	57	**63**	**144**	**96**	*45*	50	63
6	**109**	*111*	52	**117**	**121**	**129**	**150**	**83**	**94**	**133**	**79**	**115**	**145**	**80**	**120**	120	204	**150**	128	220	117	**150**	**140**	**153**	**167**	**120**	**128**	67	**115**	**63**	**144**	0	*46*	*125*	**126**
7	126	145	52	58	**130**	**129**	**150**	167	**96**	**133**	134	**115**	**145**	**80**	**120**	110	**101**	**148**	133	**109**	128	**150**	34	*100*	333	60	**129**	92	**117**	175	*94*	**93**	*46*	*125*	252
8	**108**	**81**	**108**	58	**123**	**128**	**150**	165	**94**	**133**	**77**	**115**	**146**	**80**	**120**	92	**101**	123	**66**	**109**	88	76	70	**150**	90	75	0	*77*	**112**	**63**	*95*	**94**	67	**99**	**128**
9	**109**	136	**105**	**117**	**122**	**131**	75	**83**	**95**	66	**77**	*77*	**146**	**80**	**120**	114	127	**150**	**65**	73	110	*120*	70	*51*	83	**121**	**129**	**112**	**114**	**63**	*47*	**94**	*45*	**100**	63
10	62	123	53	**117**	**126**	**129**	**150**	**82**	**94**	67	*99*	**114**	**146**	**80**	61	108	110	**144**	**65**	157	135	*60*	**142**	**153**	83	**120**	**128**	55	105	142	**145**	*141*	*45*	**102**	**126**
11	*74*	*56*	52	**117**	**125**	**128**	**150**	**83**	**95**	**134**	80	**115**	**146**	40	**120**	59	**102**	75	43	**109**	63	*59*	79	*51*	**165**	64	**131**	55	*38*	**63**	75	46	*45*	109	**125**
12	*77*	*109*	*157*	58	**123**	**129**	**150**	**83**	**94**	**133**	112	*76*	**146**	**80**	**120**	74	**102**	75	130	**111**	57	**150**	71	*101*	83	**121**	*80*	**114**	225	**64**	*96*	**94**	**134**	150	**128**
13	**115**	*111*	**105**	59	**120**	**129**	75	**83**	**95**	**133**	**80**	**115**	**146**	**80**	**120**	126	**102**	**150**	**65**	**109**	91	**150**	**144**	**154**	**167**	60	*87*	**109**	**114**	*95*	*96*	**94**	**136**	**99**	**128**
14	103	**82**	**103**	59	**125**	**128**	77	**83**	**95**	**133**	*97*	**114**	**144**	**80**	**120**	96	**101**	110	**66**	**109**	93	*60*	**144**	**150**	**167**	**120**	**128**	105	**111**	**62**	**144**	*70*	*46*	**97**	**126**
15	**114**	*111*	*157*	**117**	**125**	**129**	**152**	**83**	**94**	**133**	*98*	**117**	**146**	**80**	**120**	108	207	**148**	133	**111**	122	**150**	**143**	**152**	**167**	**122**	*85*	**111**	**116**	128	**144**	*142*	**136**	*125*	**128**
16	**111**	**82**	**103**	58	**117**	**131**	**150**	**82**	**95**	**132**	**80**	**115**	72	**80**	**120**	88	**101**	*99*	**66**	**110**	43	74	**140**	77	83	60	**131**	*77*	**112**	**63**	72	120	98	**100**	63
17	**117**	*113*	**103**	**115**	**125**	**128**	**152**	167	**95**	**133**	**78**	**116**	**146**	161	**120**	154	**102**	108	130	**109**	123	**150**	**140**	*102*	**167**	**120**	**129**	74	**115**	**64**	**144**	**95**	**136**	**100**	**126**
18	**111**	56	117	59	**122**	**131**	75	167	**95**	66	94	**116**	**144**	**80**	**120**	111	215	74	84	59	91	140	71	*52*	**167**	**121**	*87*	69	**115**	**63**	*98*	**94**	86	102	**126**
19	56	**83**	*157*	58	**122**	**132**	**154**	**83**	**94**	**131**	**79**	**114**	**146**	159	**120**	99	**99**	114	**65**	**109**	61	122	120	**146**	**167**	62	**131**	**112**	**110**	*190*	**144**	96	95	**98**	**127**
20	124	133	*157*	**119**	**124**	**128**	**150**	164	**94**	**133**	154	**116**	**147**	159	**120**	130	**101**	**146**	0	**109**	73	**150**	**144**	**157**	**166**	**120**	**130**	**114**	**116**	**64**	**144**	*142*	**134**	**101**	**126**
21	95	**83**	*157*	59	**123**	**129**	75	**83**	**94**	66	*99*	**116**	**146**	**80**	**122**	122	**104**	**144**	**65**	**109**	135	74	71	**148**	85	**120**	**128**	128	**114**	**63**	72	*71*	**136**	**99**	**126**
22	220	134	*153*	**116**	**126**	**129**	**150**	**83**	**94**	**133**	92	**114**	**146**	**80**	**120**	59	**101**	207	133	**109**	99	**149**	70	*102*	**169**	63	*86*	84	*77*	**63**	**144**	**94**	**136**	191	**125**
23	**118**	**83**	**105**	**123**	**123**	**128**	**154**	**82**	**95**	**134**	*98*	**116**	**146**	**81**	**122**	68	**100**	129	131	**109**	112	**146**	**140**	*101*	148	**125**	**131**	120	**114**	*96*	**144**	*140*	**136**	*126*	**126**
24	**117**	**81**	*157*	59	**123**	**128**	**150**	164	**95**	**134**	154	**115**	**146**	**81**	**120**	97	**103**	107	134	**109**	101	*283*	**144**	139	**164**	**121**	**129**	*81*	**117**	*97*	**142**	**94**	140	**98**	**128**
25	**120**	*111*	*157*	**119**	**122**	**128**	**150**	162	**94**	**133**	**80**	**116**	**143**	162	**125**	92	**101**	92	131	**108**	102	**152**	**140**	**150**	**165**	61	**129**	87	**112**	*195*	**144**	*70*	**136**	106	287
tempo	**114**	**82**	**105**	**117**	**123**	**129**	**150**	**83**	**95**	**133**	**79**	**115**	**146**	**80**	**120**	**?**	**101**	**148**	**65**	**109**	**?**	**150**	**142**	**150**	**167**	**121**	**129**	**112**	**114**	**63**	**144**	**94**	**136**	**99**	**126**
identical	**15**	**10**	**8**	**13**	**25**	**23**	**19**	**17**	**24**	**18**	**8**	**21**	**22**	**19**	**23**		**19**	**8**	**11**	**21**		**11**	**12**	**11**	**14**	**14**	**16**	**6**	**18**	**15**	**14**	**10**	**9**	**13**	**19**
octave	**3**		**7**	**12**		**1**	**6**	**8**	**1**	**7**	**4**	**1**	**3**	**6**	**2**		**2**	**3**	**10**	**2**		**5**	**11**	**2**	**8**	**9**	**2**	**4**	**2**	**1**	**3**	**1**	**2**	**4**	**5**
related	**2**	**6**	**8**								**5**	**2**						**1**	**1**	**1**		**3**		**6**		**1**	**5**	**3**	**2**	**4**	**5**	**6**	**8**	**4**	
rel oct		**3**	**1**									**1**						**1**		**1**		**3**		**5**					**1**	**3**	**1**	**4**	**2**		
unrelated	**5**	**6**									**2**					**25**	**2**	**12**			**25**		**1**	**1**	**3**		**1**	**12**	**2**	**2**			**1**	**2**	**1**
pattern			**1**			**1**					**6**						**2**		**2**			**3**	**1**								**2**	**3**	**3**	**2**	
meter	1	3	2	4	2	4	6	4	4	3	5	3	4	4	4	u	4	4	3	2	u	5	4	3	u	3	2	6	3	4	2	6	3	2	2
2nd meter		4	3									2						3		3				2			3		2	3	3	4			

(*Continued*)

Table A1. (*Continued*).

	T36	T37	T38	T39	T40	T41	T42	T43	T44	T45	T46	T47	T48	T49	T50	T51	T52	T53	T54	T55	T56	T57	T58	T59	T60	T61	T62	T63	T64	T65	T66	T67	T68	T69	T70
1	70	75	**105**	72	**93**	70	**82**	**91**	63	78	**122**	*73*	**94**	*44*	59	57	**81**	**91**	59	65	**125**	**58**	80	**57**	62	**98**	69	**59**	90	**93**	**109**	86	39	*70*	**41**
2	*92*	35	*69*	71	*60*	115	**81**	46	63	80	103	*74*	**94**	88	**117**	**114**	41	**92**	59	65	**125**	**58**	81	**59**	**122**	**96**	118	**59**	91	**93**	54	**154**	**79**	**92**	**41**
3	**136**	**139**	96	72	**101**	**142**	162	**92**	**128**	**157**	**122**	**108**	**94**	131	**117**	**115**	164	**92**	**117**	**128**	**128**	**58**	**162**	**57**	**121**	*133*	**131**	*88*	90	**92**	**108**	**159**	**79**	**92**	**41**
4	**136**	*52*	**104**	72	49	69	**81**	**90**	**128**	70	**122**	54	**94**	37	59	57	**81**	**91**	**119**	**128**	**128**	**59**	**159**	*74*	**122**	**98**	67	**59**	**180**	**92**	**109**	**150**	*60*	*140*	**41**
5	**140**	**140**	*70*	**144**	**94**	70	**81**	45	**128**	**159**	*81*	**107**	**95**	28	**117**	58	**83**	181	59	**131**	**125**	**58**	**163**	**58**	**123**	*131*	**133**	**61**	**180**	**92**	**109**	*103*	**79**	*140*	**41**
6	**138**	**140**	*140*	**146**	*32*	**140**	163	183	**127**	**161**	**120**	*150*	*142*	102	**118**	**115**	**81**	181	**119**	**128**	**128**	**58**	**164**	*22*	**122**	199	**136**	**60**	**178**	186	220	**154**	*117*	*140*	**41**
7	**138**	**145**	215	**144**	115	**140**	164	181	149	154	**119**	167	**97**	**111**	**117**	**117**	*99*	185	128	**129**	**131**	116	**162**	*78*	**125**	*129*	**133**	**59**	**181**	136	**111**	*211*	**79**	**94**	**42**
8	*94*	**140**	*71*	73	*62*	78	**81**	**88**	**126**	79	**120**	**108**	**92**	*44*	**117**	**114**	**81**	**92**	*91*	65	**128**	**58**	**162**	**62**	**122**	**98**	**128**	102	**181**	**92**	**109**	**154**	**79**	120	**42**
9	**140**	**140**	*70*	48	*62*	70	**82**	**91**	**129**	80	**120**	*162*	*142*	*44*	**117**	57	55	**92**	60	**128**	**126**	**57**	**164**	*19*	**122**	*131*	**136**	**59**	90	**92**	**109**	**154**	39	*138*	**42**
10	**136**	**140**	*123*	73	*65*	150	164	**91**	**128**	100	**120**	*152*	**94**	127	**117**	**115**	**82**	**95**	**117**	**131**	**126**	114	**162**	**61**	61	**99**	**136**	**59**	**181**	**94**	**109**	34	240	**92**	**43**
11	69	34	**104**	72	47	*46*	**81**	**89**	63	**162**	**121**	**108**	47	88	**117**	58	**80**	181	59	**128**	**128**	**58**	**162**	**61**	**122**	*132*	40	**59**	136	115	**110**	77	*117*	*140*	**43**
12	*92*	**138**	*70*	**145**	**95**	70	**81**	**91**	**126**	150	**122**	**109**	**96**	**117**	**117**	**114**	**81**	188	**117**	**128**	**125**	**58**	82	**62**	**122**	**98**	**138**	177	91	188	**108**	*207*	*117*	**92**	82
13	**138**	**140**	**104**	**146**	*63*	109	162	**92**	**125**	**162**	**122**	**109**	**96**	*88*	**119**	**115**	**82**	185	**118**	**128**	**125**	**58**	**164**	**58**	**123**	195	**136**	118	**181**	**92**	**108**	*103*	**79**	**93**	82
14	**138**	69	**104**	73	91	*49*	**82**	**90**	**126**	*109*	**120**	**109**	116	**107**	**119**	**114**	*64*	162	**117**	**130**	**125**	**57**	154	45	**122**	**98**	*94*	**59**	**178**	**91**	**107**	**154**	**78**	*140*	82
15	**139**	**138**	**104**	**146**	**94**	**141**	**81**	**90**	**128**	148	**120**	**107**	*144*	**111**	**117**	**115**	167	181	**117**	**129**	**128**	115	**164**	114	**123**	*131*	**134**	117	**181**	125	**108**	**152**	**79**	**92**	83
16	**140**	69	*69*	72	*63*	73	**82**	**91**	61	*104*	**120**	**109**	**92**	111	59	108	**82**	**91**	**122**	**128**	**125**	**59**	150	**57**	**122**	**98**	**136**	*89*	**178**	80	**110**	76	*59*	**92**	86
17	**136**	**136**	*140*	**146**	*128*	**136**	164	**91**	**126**	**159**	**120**	**108**	**94**	**109**	**117**	**115**	**81**	**93**	**117**	**128**	**126**	115	**162**	114	**122**	*133*	*95*	175	91	188	**109**	*104*	**79**	**92**	124
18	**136**	**142**	**103**	**146**	*62*	112	**82**	*60*	**125**	**161**	**120**	215	**94**	91	**117**	57	**82**	**93**	**119**	**128**	**126**	114	**163**	123	**125**	*129*	69	**59**	91	185	70	203	**78**	*138*	125
19	**138**	**139**	*139*	**146**	*124*	**141**	162	181	**128**	149	**120**	**108**	*140*	215	**120**	**115**	**82**	*136*	**117**	**128**	**125**	112	**164**	128	**125**	**100**	**134**	180	**181**	186	**109**	116	*117*	*140*	125
20	**137**	**140**	**105**	**146**	**100**	**140**	164	181	**128**	*211*	**120**	*162*	**95**	178	**117**	**117**	164	185	**122**	**131**	**125**	116	**162**	133	**123**	**98**	**136**	177	**181**	185	215	**157**	**77**	185	125
21	**136**	69	**104**	**144**	54	133	**81**	**91**	65	**162**	**120**	**109**	**95**	123	**119**	**115**	**82**	**92**	**120**	**128**	**127**	**57**	**159**	117	61	**98**	**129**	*88*	91	**93**	**108**	*103*	*117*	*71*	126
22	**136**	*100*	**104**	**144**	*61*	117	162	**91**	**125**	**162**	*80*	*162*	**92**	97	**117**	**116**	**81**	**92**	125	**131**	**125**	**58**	**167**	*78*	**122**	**99**	97	178	**181**	188	**109**	191	**77**	**93**	126
23	**138**	**140**	**103**	**145**	**101**	**139**	**82**	**91**	**126**	*107*	**120**	**108**	**94**	**110**	**119**	**117**	**82**	*136*	**112**	**131**	**125**	116	151	*81*	**125**	*131*	**140**	*88*	91	**92**	**107**	138	*119*	*140*	126
24	**136**	**140**	**103**	**145**	**99**	153	162	**90**	**128**	*106*	**120**	**109**	**93**	**118**	**118**	172	167	175	**117**	**129**	**125**	**59**	**162**	129	**124**	*131*	117	178	**181**	**91**	**109**	*101*	*117*	*140*	136
25	**139**	**140**	**104**	**142**	**101**	*98*	**81**	**90**	140	95	138	**109**	*140*	**117**	**120**	**116**	162	193	**114**	**131**	**129**	**56**	101	128	**125**	**102**	**136**	**59**	89	**91**	**105**	92	**80**	**94**	141
tempo	**138**	**140**	**104**	**145**	**99**	**140**	**81**	**91**	**127**	**161**	**120**	**108**	**94**	**111**	**117**	**114**	**82**	**92**	**117**	**129**	**126**	**58**	**162**	**59**	**123**	**98**	**136**	**59**	**180**	**92**	**109**	**154**	**79**	**92**	**42**
identical	**20**	**17**	**13**	**15**	**9**	**8**	**15**	**18**	**18**	**9**	**21**	**15**	**18**	**8**	**22**	**17**	**16**	**12**	**17**	**22**	**25**	**17**	**18**	**10**	**22**	**13**	**16**	**12**	**14**	**14**	**21**	**9**	**13**	**11**	**11**
octave	**2**	**6**	**1**	**9**	**2**	**6**	**10**	**6**	**5**	**4**		**2**	**1**	**2**	**3**	**6**	**6**	**8**	**5**	**3**		**8**	**3**	**8**	**3**	**2**	**3**	**8**	**10**	**6**	**3**	**3**	**3**	**1**	**11**
related	**3**	**1**	**6**	**1**	**8**	**2**		**1**		**4**	**2**	**5**	**5**	**3**			**1**	**2**	**1**					**4**		**10**	**2**	**4**	**1**		**1**	**5**	**7**	**10**	
rel oct		**1**	**4**		**3**	**1**				**1**		**2**		**1**		**1**	**1**							**2**								**2**	**2**	**2**	
unrelated			**1**		**3**	**5**			**2**	**7**	**1**	**1**		**5**		**1**	**1**	**3**	**2**				**4**	**1**			**2**	**1**		**4**		**5**		**1**	**2**
pattern						**3**					**1**		**1**	**6**													**2**			**1**		**1**			**1**
meter	3	4	3	6	3	5	2	2	2	3	5	2	6	5	4	2	5	2	4	4	4	4	1	3	1	3	6	6	2	6	6	4	6	4	2
2nd meter	4		4		2					4	4	3	2					3	3									2				3	3	6	

Appendix B: Beat tracker annotations

Table B1. Tempo from a set of beat tracker and tempo estimation tools, * by Zapata.

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18	T19	T20	T21	T22	T23	T24	T25	T26	T27	T28	T29	T30	T31	T32	T33	T34	T35
IBT*	215	**82**	52	**116**	*98*	258	75	167	**96**	66	*99*	**115**	73	161	**120**	235	207	75	**65**	215	152	**152**	52	102	**167**	*81*	258	*167*	92	126	**144**	**94**	**136**	**99**	63
Audiosculpt	**118**	164	*156*	**117**	102	**129**	**149**	164	**95**	**132**	**79**	**115**	**145**	161	**120**	80	**101**	76	**65**	**109**	139	*120*	**141**	102	83	**121**	**129**	*167*	**114**	126	**143**	**93**	**136**	**100**	**126**
beatCounter106	68	202	*78*	89	*185*	64	75	165	**96**	*171*	126	**115**	*108*	160	240	51	**101**	99	54	218	89	*90*	*196*	*51*	83	**121**	199	56	230	*95*	**143**	**93**	68	**99**	**126**
Böck	230	176	*157*	**115**	**125**	240	**150**	166	**95**	**133**	*98*	230	72	162	**120**	146	200	222	*86*	**109**	230	**153**	*92*	101	176	**122**	**130**	133	230	127	*95*	**93**	**136**	**100**	**125**
Stark-Davies	**117**	**81**	*156*	**119**	104	**129**	**150**	**83**	**95**	**132**	*106*	*76*	**144**	*102*	**120**	101	**101**	**141**	*89*	**110**	94	*60*	132	101	**166**	*78*	**129**	**113**	136	125	*95*	**93**	*90*	**97**	**126**
Klapuri*	**112**	**81**	*157*	**117**	**126**	**129**	**150**	**83**	**95**	**133**	*99*	**115**	**146**	**80**	**120**	109	**101**	**148**	*87*	**109**	115	*120*	*94*	102	83	**122**	**129**	**112**	**115**	126	*95*	**93**	*91*	**100**	**126**
ILSP	*75*	**80**	*78*	**117**	**124**	**129**	75	165	**95**	**132**	*99*	**115**	**145**	160	**120**	143	203	213	129	217	197	*60*	**140**	101	176	**120**	**129**	144	230	127	**143**	**94**	**136**	**99**	**126**
aufTAKT	**112**	**80**	*78*	**118**	**124**	**130**	75	**83**	**94**	66	*98*	**116**	*110*	160	90	70	88	136	*88*	**109**	165	*90*	**141**	**153**	83	**121**	**128**	**112**	**115**	128	**142**	**94**	58	**102**	**127**
Ellis	**116**	*111*	*155*	**118**	**125**	**128**	**148**	*111*	**96**	**131**	*100*	**115**	**144**	*108*	**120**	93	135	**140**	128	**109**	129	**153**	**147**	100	**161**	**124**	**128**	136	*153*	128	*96*	**94**	**137**	**100**	**126**
MixMeister	**112**	134	*156*	93	*95*	**128**	**149**	**82**	**94**	**135**	*100*	123	**144**	160	**120**	**128**	97	117	130	**108**	145	**150**	**141**	99	95	**120**	**128**	104	96	125	*95*	**94**	*91*	*152*	**126**
Aubio	129	*108*	*161*	128	**126**	**129**	**150**	**83**	**95**	**127**	131	**115**	**145**	115	**120**	**128**	**101**	**141**	*88*	**109**	88	*60*	127	102	**169**	**121**	**128**	128	*154*	127	**143**	**94**	**135**	130	**125**
BtRule	86	167	*157*	**116**	**121**	*86*	**150**	165	**95**	**133**	*197*	*153*	**144**	160	178	**122**	**101**	128	130	**109**	111	**150**	105	**152**	66	**120**	*86*	*165*	**114**	*94*	**143**	*140*	**136**	**100**	**126**
Degara*	133	**81**	*157*	**115**	103	**129**	**152**	**82**	**94**	**133**	**78**	**115**	**144**	*108*	**120**	**129**	133	**148**	*86*	**108**	101	*60*	**140**	101	**167**	**120**	**129**	**112**	*152*	126	*96*	**94**	*91*	**99**	**126**
Ellis*	231	188	*156*	**117**	**123**	259	**150**	326	**95**	**133**	*99*	231	**146**	319	**119**	92	203	**142**	259	217	300	**153**	*91*	**152**	**167**	242	259	366	*153*	127	283	**94**	**136**	246	250
Beatroot*	126	183	*154*	185	171	**129**	**150**	164	**94**	**133**	*194*	**113**	**146**	160	**120**	150	80	**150**	130	**109**	182	**150**	**141**	103	**167**	**121**	**130**	**113**	*153*	128	**146**	**94**	**136**	**100**	**125**
MIRtoolbox	**115**	161	*156*	186	63	**128**	**149**	*110*	*124*	**128**	*95*	**116**	**144**	161	**120**	55	58	89	*88*	**109**	148	*60*	56	100	**168**	**122**	*86*	95	*153*	126	*96*	**94**	**137**	**100**	**126**
Sonic Annotator	66	**83**	*52*	**117**	107	**132**	**152**	**83**	**96**	**136**	*105*	**117**	**148**	**81**	**123**	**129**	58	76	*88*	**110**	103	*60*	**144**	103	**167**	**123**	**132**	89	*157*	129	*97*	**96**	**136**	**99**	**129**
Tempo (human)	**114**	**82**	**105**	**117**	**123**	**129**	**150**	**83**	**95**	**133**	**79**	**115**	**146**	**80**	**120**	**?**	**101**	**148**	**65**	**109**	**?**	**150**	**142**	**150**	**167**	**121**	**129**	**112**	**114**	**63**	**144**	**94**	**136**	**99**	**126**
BT = Human	1	1	0	1	1	1	1	0	1	1	0	1	1	0	1	0	0	1	0	1	0	1	1	0	1	1	1	0	0	1	1	1	1	1	1
identical	7	7	0	12	8	12	13	7	16	14	2	13	13	2	14		6	7	2	13		7	8	3	9	14	12	5	4	0	8	16	11	14	15
octave	2	3	2		1	3	4	8		2		2	2	11	1		4	3	6	4				1	4	1	2	2	3	15	1		1		2
related	1	2	12	2	7	1		2	1	1	11	1	2	3	1			2	8			6	3	13		2	2	3	7	2	8	1	4	1	
rel oct			3	2	1						2	1			1							4	1												
unrelated	7	5		1		1					2			1			7	5	1				5		4		1	7	3				1	2	

(*Continued*)

Table B1. (*Continued*)

	T36	T37	T38	T39	T40	T41	T42	T43	T44	T45	T46	T47	T48	T49	T50	T51	T52	T53	T54	T55	T56	T57	T58	T59	T60	T61	T62	T63	T64	T65	T66	T67	T68	T69	T70
IBT*	*91*	*103*	*69*	72	207	**140**	**81**	178	258	**161**	246	**108**	93	**108**	235	57	*162*	**92**	**120**	**129**	**126**	**57**	**167**	*161*	246	*133*	258	117	**185**	184	**108**	162	*117*	184	82
Audiosculpt	**138**	*207*	*138*	**144**	**92**	**141**	164	**91**	102	**159**	**121**	**108**	94	**109**	**119**	**114**	**82**	184	**118**	**129**	**126**	114	**164**	*77*	**122**	*131*	**130**	175	90	*124*	**108**	**154**	**79**	**93**	83
beatCounter106	*92*	**138**	*69*	72	109	*58*	**81**	**91**	**127**	**160**	**122**	55	93	170	**119**	*86*	**82**	**92**	104	**127**	**126**	*85*	**163**	*77*	**122**	196	*92*	*88*	89	*119*	**109**	**155**	**78**	184	83
Böck	*92*	68	*69*	**146**	240	92	162	181	**127**	**162**	**122**	**109**	93	73	240	*157*	*162*	181	162	**130**	166	*187*	**162**	*150*	**122**	**98**	*90*	171	**181**	187	222	*101*	**77**	**92**	84
Stark-Davies	*92*	**138**	*139*	**144**	**99**	*114*	**81**	**90**	116	*105*	**117**	**107**	93	95	**119**	**113**	*163*	**91**	**115**	**128**	**125**	115	82	*77*	**124**	**99**	*93*	117	90	*125*	**107**	*104*	**79**	**94**	84
Klapuri*	*91*	**138**	*69*	**144**	**93**	**140**	**81**	**91**	**126**	*107*	**120**	**109**	94	**109**	**117**	**114**	**81**	**92**	**117**	**129**	**128**	115	107	*80*	**122**	**102**	*91*	117	91	*123*	**108**	*103*	*117*	**92**	85
ILSP	*91*	*207*	*69*	**143**	**92**	**142**	**81**	**91**	*84*	**160**	**120**	**108**	94	**110**	**118**	**114**	**81**	**92**	**117**	**130**	**127**	114	**164**	*161*	**122**	**98**	*91*	118	90	*124*	217	*103*	**78**	**93**	85
aufTAKT	*91*	69	**103**	*94*	*138*	*83*	**82**	**90**	**121**	**159**	**118**	**106**	95	**109**	**117**	**114**	*163*	**93**	144	**135**	154	115	82	*77*	**122**	**98**	*94*	145	91	**92**	**108**	**155**	*117*	**93**	85
Ellis	**132**	**138**	**101**	**144**	**94**	**139**	158	*123*	**127**	*107*	**122**	**109**	94	**110**	**119**	**114**	*163*	111	**115**	**128**	**125**	118	**163**	*155*	**124**	*130*	**129**	120	*134*	*125*	*142*	*102*	*120*	**91**	125
MixMeister	*93*	*103*	**103**	**143**	**94**	*112*	162	133	**125**	**159**	**121**	**104**	93	**108**	**118**	**114**	*154*	**91**	**117**	**127**	**125**	95	**162**	97	**123**	**100**	*95*	131	90	*124*	91	123	*127*	**94**	127
Aubio	*92*	69	*35*	**144**	**96**	**144**	41	181	**128**	**160**	*85*	**108**	93	128	**119**	57	*162*	**91**	162	**127**	**123**	117	*126*	129	**122**	*131*	*96*	**59**	96	*124*	*144*	*103*	**80**	**93**	129
BtRule	**138**	*207*	**104**	72	**96**	**141**	**81**	*120*	*167*	**160**	**120**	**107**	94	**110**	**118**	**114**	**82**	138	**119**	**127**	**125**	115	**163**	*78*	**122**	196	**133**	176	91	186	*144*	**154**	*117*	**93**	162
Degara*	*92*	**138**	*69*	**144**	103	*112*	161	185	*167*	*108*	**120**	**108**	94	*86*	**117**	**115**	*161*	**92**	**117**	**129**	**126**	115	**161**	144	**123**	*129*	*90*	117	**178**	*123*	**108**	*103*	*117*	**92**	167
Ellis*	*183*	**138**	238	**144**	186	**142**	163	*366*	254	319	242	**109**	94	**109**	234	*155*	*163*	273	221	**129**	169	224	**165**	*155*	246	*133*	**144**	341	**181**	*124*	217	*207*	234	185	167
Beatroot*	*91*	**136**	*141*	**146**	188	**143**	164	182	**128**	**160**	**120**	**109**	94	154	**118**	**114**	*162*	182	**119**	**130**	**126**	114	**162**	180	**122**	*130*	*91*	176	**182**	187	*141*	**154**	*118*	187	169
MIRtoolbox	*189*	**139**	*69*	**147**	183	94	161	*123*	**126**	**160**	*86*	**107**	84	119	**118**	*156*	*163*	183	**120**	**128**	167	*84*	**164**	*152*	**123**	**99**	*94*	72	**179**	*124*	**107**	*103*	**78**	*133*	170
Sonic Annotator	*93*	**140**	*70*	**148**	**96**	*56*	167	**92**	**129**	**162**	*85*	**108**	94	**112**	**120**	**115**	*167*	**92**	**120**	**129**	**126**	129	**167**	147	**123**	*133*	*96*	117	91	*126*	**110**	*103*	*120*	**94**	172
Tempo (human)	**138**	**140**	**104**	**145**	**99**	**140**	**81**	**91**	**127**	**161**	**120**	**108**	**94**	**111**	**117**	**114**	**82**	**92**	**117**	**129**	**126**	**58**	**162**	**59**	**123**	**98**	**136**	**59**	**180**	**92**	**109**	**154**	**79**	**92**	**42**
BT = Human	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1	0	0	0	0	0	1	0	0	1	0
identical	3	9	4	13	9	9	7	7	10	12	12	16	17	11	14	11	5	10	12	17	13	1	12	0	15	7	4	1	6	1	9	5	7	12	0
octave		3		3	3		10	5	2	1	2	1			3	2	12	5	1			11	3		2	2	1	12	10	4	3		1	4	17
related	12	3	8	1	1	3		3	2	4	3			1		4					3	2		6		8	12	1	1	12	4	9	9	1	
rel oct	2	2	4			3		1	1													1	1	6								1			
unrelated			1		4	2		1	2					5				2	4		1	2	1	5				3			1	2			

Appendix C: List of digitization efforts

TAble C1. Digitization efforts of music collections with, at least some, ethnic music.

Acronym	Name	Coordinator	URL
DEKKMMA	Digitization of the Ethnomusicological Sound Archive of the Royal Museum for Central Africa	IPEM, Gent University	http://music.africamuseum.be
DELOS	Network of Excellence on Digital Libraries	Association	http://www.delos.info
DISMARC	Discovering Music Archives Across Europe	Consortium	http://www.dismarc.org
EASAIER	Enabling access to sound archives integration enrichment retrieval	Queen Mary University, London	http://www.elec.qmul.ac.uk/easaier
EthnoArc	Linked European Ethnomusicological Archives	Wissenschaftskolleg zu Berlin	http://www.ethnoarc.org
Kopal	Co-operative development of a long-term digital information archive	Niedersaechsische Staats- und Universitaetsbibliothek Goettingen	http://kopal.langzeitarchivierung.de
Nestor	Network of Expertise in Long-Term Storage of Digital Resources	Deutsche Nationalbibliothek, Frankfurt am Main	http://www.langzeitarchivierung.de
MICHAEL	Multilingual Inventory of Cultural Heritage in Europe	EU countries	http://www.michael-culture.org
POFADEAM	Preservation and On-line Fruition of the Audio Documents from the European Archives of etnic Music	University of Udine	http://www.ipem.ugent.be/?q=node/19
PrestoSpace	Preservation towards storage and access. Standardized Practices for Audiovisual Contents in Europe.	Institut National de l'audiovisuel, Paris	http://www.prestospace.org
TAPE	Training for Audiovisual Preservation in Europe	European Commission on Preservation and Access (ECPA)	http://www.tape-online.net

Cornelis, O. (2013). From information to inspiration, sensitivities mapped in a casus of Central-African music analysis and contemporary music composition. Critical Arts. 27 (5) pp. 624-635.

From information to inspiration, sensitivities mapped in a casus of Central-African music analysis and contemporary music composition

Olmo Cornelis

Abstract

The increased digitisation of cultural objects has resulted in a vast resource of accessible research data. For researchers (and artist-researchers) this abundance of information requires critical analysis and assessment of the available data, and calls for new kinds of scientific analysis. In the case of ethnic music, this analytical framework for digitised and digital-born audio objects is offered by computational ethnomusicology. This article focuses on the position of a composer-musicologist (the author) who works with digitised Central African music on a daily basis, which inevitably influences his musical idiom. First, a historical overview of ethnomusicology is drawn which confronts two strands of research within that field that can be related to elements from Hal Foster's article 'The artist as ethnographer?' (1995). Second, a brief outline is given of the artist's artistic and scientific research, and how they are related. Finally, the discussed aspects of ethnomusicology, the research and composition, are considered in light of Foster's thesis.

Keywords: artistic research, Central Africa, composition, computational ethnomusicology, Hal Foster, music

Introduction: a historical overview of ethnomusicology

An historical overview of the study of non-European music reveals an interesting dichotomy that is still of relevance to the contemporary artist with a multicultural interest.

Olmo Cornelis studied musicology (Ghent University) and composition (University College Ghent) and is currently writing a PhD at the School of Arts, University College Ghent. olmo.cornelis@hogent.be

critical arts
Routledge Taylor & Francis Group
UNISA university of south africa PRESS
ISSN 0256-0046/Online 1992-6049
27 (5) 2013
pp. 595–605
DOI: 10.1080/02560046.2013.855524

While the oldest writings on non-European music date from the 17th century, as for example the work of Praetorius (1619), Mersenne (1636) and Dapper (1668), the more systematical approaches to the description of non-European music can be found from the 18th century onwards, especially with authors such as Joseph Amiot and Jean-Jacques Rousseau. Rousseau's *Dictionnaire de musique* (1768) contains transcriptions and descriptions of European folk music, Chinese music, and native North American music. Because of its emphasis on comparing cultures and highlighting differences between them, this dictionary can be seen as the start of 'comparative musicology' – a term coined by Guido Adler in 1885. Its basic principle was to study and understand music by comparing musical expressions of different primitive cultures that were commonly seen as static cultural entities.

There was, however, an important discrepancy between the approaches of German and American scholars. The German school, with Carl Stumpf, Curt Sachs and Erich von Hornbostel, was characterised by acousticians and psychologists. The American scholars – not organised into a single school, as such – were primarily anthropologists and musicians. The former group focused on the study and understanding of music by comparing different cultures; their overall aim was to find 'universal' features of music. The latter group researched the place of music *within* a culture: its functions, meanings and origins. The differences between German and American scholars can also be explained from a geographical perspective: German researchers depended on archived objects and scarce documentation that arrived from overseas in those days. There was little access to the culture being studied, which meant the available recordings were inevitably analysed in exhaustive depth. In practice, these scholars displayed great interest in the analysis of pitch phenomena (e.g., melody, scales, intervals) – the most tangible parameters to analyse. The American researchers, however, were much more closely connected to their fields of study, of native Native American music. Given their proximity, it was possible to engage ethnographically with the native cultures and to conduct fieldwork that focused on societal and contextual elements.

After the World War II, the term 'comparative musicology' declined radically in use as its methodology of comparing musical cultures was no longer accepted. Around 1950, Jaap Kunst introduced the neologism 'ethnomusicology' (Kunst 1950), a discipline that focuses more on the 'ethnic' part of music, proceeding from a more anthropological angle in which fieldwork becomes an essential part of research, and where contextual information is considered crucial (Nettl 1964). Ethnomusicology was seen as an interdisciplinary field of research that not only combines musicology, anthropology, sociology and folklore, but also biology, psychology, historical musicology, dance and linguistics. Bruno Nettl (ibid.) writes that ethnomusicology is therefore not a discipline, but a field in which several disciplines merge. In the following decades, a confrontation ensued between the more anthropological,

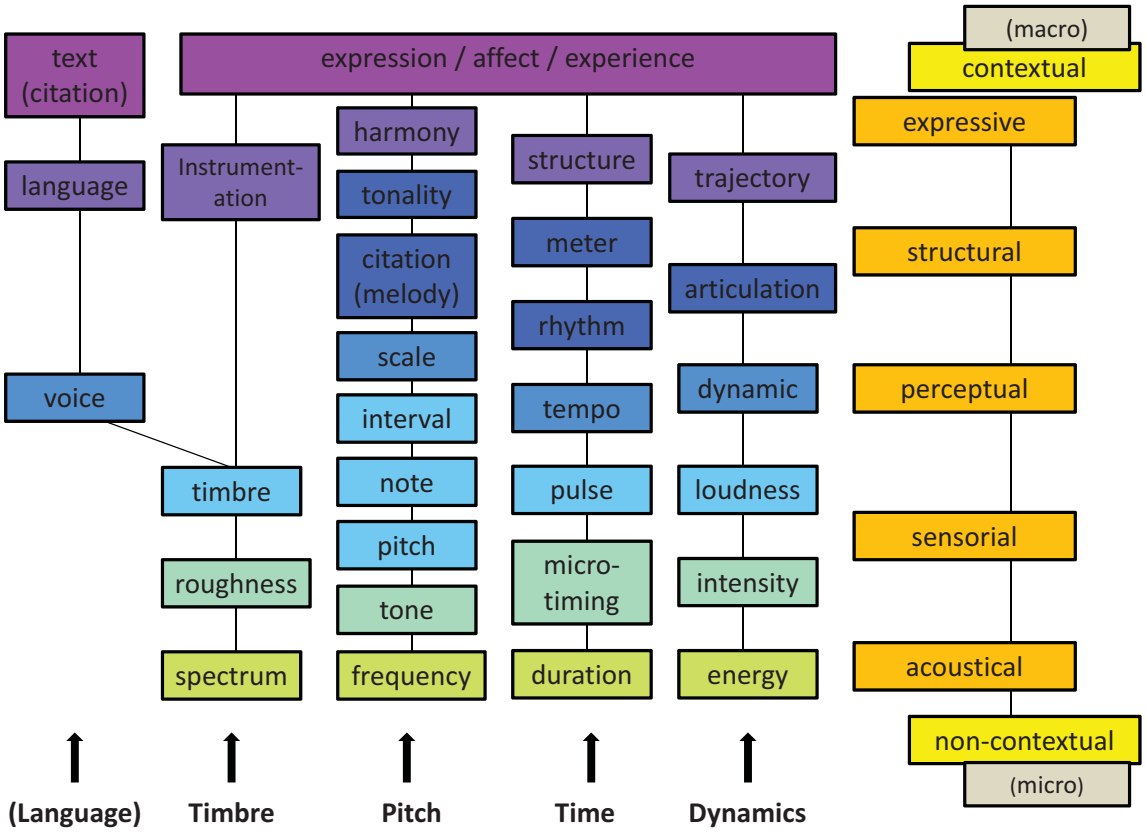
sociological approach and the more analytical, musicological approach. Curt Sachs' approach, for example, could be considered the most 'comparative' of his generation; in *History of music* (1955) he aims to find an explanation for the origin of music by studying the most primitive musical expressions in the world, hoping to create a general theory on the universality of music. Alan Merriam (1977), on the other hand, criticises the comparative German approach, and stresses the importance of social, anthropological and cultural factors, while considering fieldwork essential.

From the 1970s, this traditional division became less explicit and ethnomusicologists covered both perspectives in their work, integrating both the study of sonic elements and the cultural context in which they are produced. Ethnomusicological research has since broadened towards new focal points such as historical dimensions; gender aspects associated with musical activity; iconographic representation; motor aspects related to the production of sound and dance; the relationship between the execution of musical instruments and the human body; and many other issues (Oramas 2012).

As a result of the increasing trend towards digitising sound and sound archives, computational ethnomusicology (Cornelis 2010; Tzanetakis, Kapur, Andrew Schloss et al. 2007) has emerged as a recognisable scholastic approach since the beginning of the 21st century. Computational ethnomusicology usually employs one of three main analytical methods: signal analysis, symbolic analysis and semantic analysis. It implies a shift in methodology: computational analyses are extremely detailed and go beyond human annotation, are significantly less time-consuming and can be performed on very large sets of audio data. They reinforce methodologies that lean towards the *German* school, with systematic analysis of musical parameters now being performed on large archives. Critics of this new approach address the difficulty of a) analysing a digital signal that lacks contextual information, and b) analysing (polyphonic) music automatically. Critical human monitoring of and adjustments to automatically obtained data and descriptions, therefore remain necessary.

Methodology

The research that forms the basis for this article is a part of a PhD study in art, which combines the results of ethnomusicological analysis with artistic experiment. Being a disciple of computational ethnomusicology, I mainly use signal analysis to obtain computer-assisted descriptions of Central African music that aims to encapsulate musical characteristics. My artistic work is connected to this scientific research and to attempts to merge Western and non-Western musical elements and techniques. Each composition created during this research attempts to use a different compositional approach. Sometimes there is an explicit use of ethnic elements within my music, at other moments it is implicit and noticeable only to an



Bipolar Axes:

- Explicit <----------> Implicit
- Effect <----------> Concept
- Form <----------> Musical
- Music as Such <----------> Functional
- High level <----------> Low level MP
- Physical <----------> Virtual
- ... <----------> ...

Figure 1 (Top): Display of most musical parameters (pitch, time, timbre, dynamics) with their organisation from micro- to macrostructure, i.e., from low-level non-contextual features to high-level contextual elements (Orange displayed categorisation based on Lesaffre, Leman and Tanghe et al. [2003]).

Figure 1 (Bottom): Several bipolar axes of interest for my personal classification of musical parameters.

informed audience. My scientific research, with its computational approach, offers insights into Central African music which, in turn, influence my artistic experiments. For personal reflection and awareness as an artist using ethnic influences, I have drawn up a theoretical model that lists musical parameters along several bipolar axes. Figure 1 shows a general organisation of musical parameters, while Figure 2 shows an example of the same parameters along bipolar axes.

Technical description of recent works of the author

'From fiddle to fickle' (2008)

'From fiddle to fickle' is a short electronic composition that experiments with the possibilities of an audio sample. The piece is based on a single song from an African recording made available by the RMCA (Royal Museum for Central Africa), Belgium. The original piece was recorded in 1969 in what was Ankole and is now South Uganda, and contains fiddle, percussion and voice. The piece is interesting because the two melodies of the fiddle behave as counterpoint. 'From fiddle to fickle' begins with the original sample. After several seconds the sample is layered upon the original by which it turns into a canonic four-voiced fugue. The sample then begins

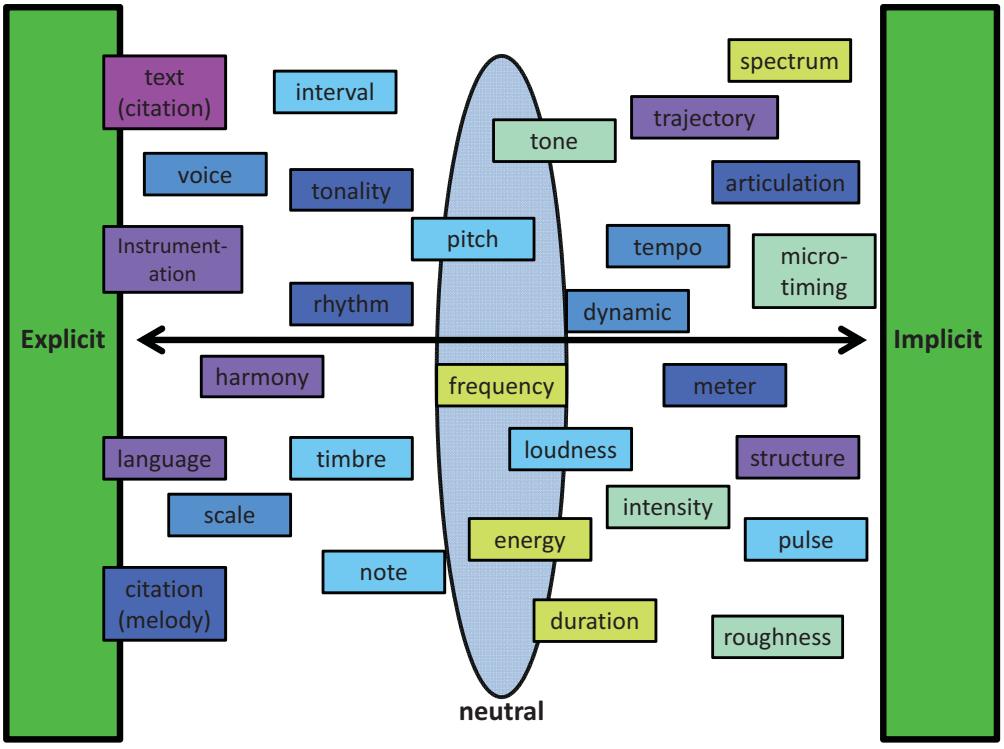


Figure 2: Example of a classification of musical elements onto a bipolar axis of explicit to implicit influence. Colours refer to the higher-level features from Figure 1. My research and composition are rather connected to the lighter blue displayed elements.

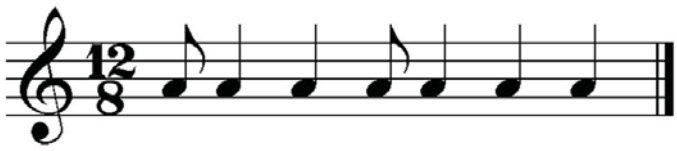


Figure 3: Famous 12/8 pattern which is played repeatedly, characterised by on/off beat shifts.

to morph, alter and deform using several electro-acoustic production techniques and effects. Gradually, the piece becomes entirely unrecognisable, ending in a totally amorphous sound that finally fades into a growl.

'Traces from Tracey' (2011)

This composition explicitly investigates the field of tension between performer and audience. In Western culture, there is a distinct difference between the two, increased by the physical border between stage and audience, where one is expected to listen quietly, and show appreciation afterwards with applause. This is in contrast to the musical expressions of oral musical cultures (Aubert 2007). In Central African music, for example, each listener may be, at the same time, a participant. The duration of a song (several hours versus a three-minute or so pop song) has an entirely different impact on the audience, and musical cadences, typical in Western music, are uncommon.

In 'Traces from Tracey' these differences in behaviour were confronted with a traditional Western setting: the audience members, unsuspectingly, entered the concert room, were compelled to find a place among the musicians, and were handed several rhythmical musical instruments. They were given a brief explanation for joining the music by playing some ostinato patterns that typically appear in Central African music. Such a repeating pattern is also called a time line, time keeper, topoi (Agawu 2006; Koetting 1970), which is, in this composition, the famous 12/8 pattern (see Figure 3). In this transcription, it is notated in a 12/8 signature, showing the entire pattern. The first part has syncopated notes off the beat, the second has notes on the beat. Note that all of the musical concepts used here are Western interpretations. In this notation one can even discuss the actual beginning of the pattern, that is often seen as circular.

'In between' (2012)

Almost all Western music relies on the equal-tempered tuning that divides the octave into 12 equal parts. However, musical scales can be organised very differently. In Arabic music, scales profile themselves by several intervals smaller than the 100 cents that typically constitute a semitone (one-twelfth of an octave) in Western music. In Central African music many different kinds of scales prevail, from equal-tempered

pentatonic to more complex divisions with mirrored intervals. In the miniature series 'In between', several different microtonal scales have been used. Each scale has its own flavour, resulting in pieces with a very diverse character, especially since the melodic and rhythmical material is in a typically Western musical idiom.

'In Cinder' (2013)

This composition is an electro-acoustic work: it combines instrumental score with a tape recording. The tape is an extreme example of sampling technique, namely concatenative composition: an audio file is segmented into very small time units, up to 100 samples per second which are then reassembled into a new audio file. In this case the CataRT software from IRCAM (Schwarz 2006) reorganises the audio segments along auditory axes, such as pitch, intensity, dynamics or spectral information. It leads to a 2D-graphical layout where all audio units with similar characteristics are clustered together. The final step is to reassemble the audio units into new audio tracks. Compositionally, one attempts to create an interesting sonic path through the graphical display and the playback options create a new audio file along the chosen settings. The results can be very surprising; sometimes offering traces of the initial sound, sometimes creating a distorted sound that does not resemble the original audio to any degree. CataRT can analyse a single audio file, though two or more files can also be arranged into one graphical display. This composition tested the technological limits of the CataRT software, arranging the African samples from the RMCA on its smallest sample rate (100 samples per second) and reassembling them with intense pitch manipulation. The result is a tape that varies from familiar, to recognisable, to vague, to mysterious, to noise.

Analysis of artistic approach, versus Foster

In this vignette, I aim to confront the historical background presented above, my own position as a researcher, and my musical compositions with some concepts from Foster's 1995 essay, 'The artist as ethnographer?'

1) As a composer-musicologist, I have created a bipolar model that organises musical elements along several axes (see above). These different axes – from explicit to implicit influence, from concept to effect, from anthropological insight to *l'art-pour-l'art* – aim to create an awareness of the actual position of the artist, especially because of the symbiosis of (scientific) analysis and personal artistic practice. Such a framework avoids the danger of *ideological patronage*, but allows informed artistic choices on an aesthetic, a functional and a conceptual level, whether purely functional for the production of an art object, or whether serving to make a statement. Any influence or reference in my music can be localised, framed and defined. The electronic component of 'In Cinder' is a fusion of short audio

excerpts. It is an explicit kaleidoscope of Central African music containing hundreds of samples that are, intentionally, used without any further notice of their content, function or origin. In this context, the re-use of (recorded) sound is a conceptual idea, and the abandonment of all its content and meta-data conflicts with the traditional anthropologist/ethno-musicological design. It deals with the effect that the rush of sampled material generates. C.L. Krumhansl (2010) argues that timbres can be recognised by leaning only on a very brief stream of information, which is also the case in 'In Cinder': somehow, people have created a platonic residue of types of African music. For very short audio fragments this is reflected in timbres of instruments and voice, language, and the specific tonal scales (intervals).

For the artist, compositions are part of a learning process, part of the path one has to follow to achieve a personal artistic expression (that can never be fully achieved). An artist with an interest in ethnicity should deal with two criteria in this matter: a full awareness by the artist of what he uses, what he does/tries to do, and what he investigates/tries to investigate. The second element is the requirement that the artist does not solely copy (out of effect), but that s/he merely integrates elements in his/her own art. The artist should strive for personal expression which is not based on copying elements, but should focus on where information nurtures inspiration, as the title of this article suggests. Plain copying or extrapolation does pertinently risk *exoticising* 'the *other*' (Foster 1995). In my work the idea of *ideological patronage* (ibid.) could well be reversed, attempting to confront a Western public with musical concepts that are unfamiliar or unknown to them: these are our traditional concepts that keep us within certain accepted borders, excluding feelings of discomfort. The composition 'Traces from Tracey' depicts such Western cultural ideological patronage, since it is the public who suddenly became 'exposed'. In this primarily sociological experiment 'we' became the 'other', since the audience was urged, by the composer, to (re)act in an unfamiliar setting. The audience members were removed from their ecological comfort zone (Leman 2007), and were hesitant about performing collaboratively with the musicians. After a few individual attempts at rhythmically joining the musicians, most members of the audience joined in, playing (nothing more than) the suggested basic rhythmic pulsation. The piece 'From Fiddle to Fickle' is another example of ideological patronage: the beginning of the composition consists of an original African song in which one could more or less track compositional techniques such as counterpoint and fugue, but these concepts are superposed by the artist in developing the composition. This could not only be seen as some kind of ideological patronage, but also as its reverse: Whose culture is being exposed? (since the typical Western compositional elements are exposed).

2) As a musicologist-composer, I noticed that the *artist as ethnographer* did encounter a new rival in the *academic artist*, with – for example, in Flanders – the recent possibility of obtaining, at art schools (in collaboration with universities), a

PhD in art: a PhD based on scientific research intertwined with artistic realisations. This could be a mode of self-presentation that indeed would foreground the emblematic figure of 'research' as one of the key problems within the ethnographic turn in contemporary art, which resolves the *artist-envy* and *ethnographer-envy*, because it serves an interdisciplinary approach gaining content and contextual information, and implies reflection, self-critique and personal adaptation within an artistic trajectory.

From this very position as a researcher-artist, however, the aspect of outsideness becomes paradoxical: on the one hand I act as a complete outsider who has not engaged in any (musical) fieldwork. On the other hand, though, my daily occupation of analysing scales and temporal features in Central African music has given me interesting insights into scales and scale evolutions (Moelants 2009), and numerous opportunities to listen to many different scales (Cornelis 2012). Musical scales were retrieved, revealed and demystified using the TARSOS software. This creates an informed artist, who relies on computational methods to enhance the tangibility of scales, and eventually to reuse them in his/her own compositions. The 'In between' series was produced in this manner. Therefore, the claim not just to speak about the other but to speak from *within*, is not located in a physical *within* per se, but, in my case, rather in a cognitive *within* and in the end, maybe even already some kind of sub-conscious *within*. It is not about a *limited* or *automatic access* by in- and outsiders, but the possibility of informed access.

3) The musicological perspective connects historical ethnomusicology to Foster's *misplaced temporalisation*, whereby non-Western practices are sometimes seen as a throwback to earlier, more primitive forms of humanity. It is remarkable that similar assumptions formed the basis of comparative musicology (see above): there was a belief that non-Western music was essentially static and unchanging (Nettl 1985). It guaranteed the validity of research that was performed on a small number of samples at hand. These theorems and methodologies were criticised and adapted when new data, and the means of transcribing such data, became available: more recordings, longer recordings, fieldwork, contextual information, video recordings. In current research, with the emergence of digital archives another shift has become noticeable: the *field came to the scholar*, and the rise of *computational ethnomusicology* with its proper approach and its potential for scanning large archives acts as a time-window.

4) As a composer, I believe cultural influences are inherent and necessary to any composer, whether it is about local or global interests, topical or foreign elements, and whether you are in an emic or etic position. Is being 'informed' and being aware of your influences a requirement for an artist? Or is the way an artist deals with this, his/her own responsibility? The cultural legacy someone wants to use, reuse or abuse (who's to judge?) has to be respected by the artist, but the artistic result is neither dependent on, nor responsible to, this legacy. Is being influenced a sign of

respect, or is it a burden you have to carry with you as an artist when it comes to accountability? Is it freeware or shareware? And do ethnic influences carry a bigger *burden* than Western influences, if Western artists use them? Within this perspective, postcolonial theories are themselves an academic discourse, as explained by Michel Foucault's (1984) 'blackmails of enlightenment'. For Foucault, a critical attitude should commit us to critically engage with history, which is a complex elaboration of our social reality and constantly occurs in a state of flux, by which a permanent (re)creation of ourselves in our autonomy is needed. This can well be transposed onto the artist. Especially in our glocalised societies, where the commercialisation of music (by [re]production, media and Internet) (Sabbe 1996) has led to an unknown level of accessibility to the world's heritage, it is apparent that an artist is flooded with cultural elements which are partly integrated in his/her very own cultural language/idiom. Our musical society has become glocalised, making concepts such as emic/etic an outdated framework, liberating us from postcolonial fears and accountabilities, but creating a large and complex web of music in all its styles and interrelations. Of interest in this perspective is Walter Wiora's book of the *Four ages of music* (Wiora and Gaudefroy-Demombynes 1963): in the first stage, music was more or less homogeneous, especially in the way cultures began to produce music. This changed in the second and third stage because of dispersed social and cultural developments. In the fourth stage, however, it all seems to reconverge due to the nature of global industrial culture. About this last stage, Nettl speaks of the intensive imposition of Western music and thoughts upon the rest of the world, and mentions the wide scope of possible responses that vary in terms of maintaining, preserving, modifying or abandoning their own musical traditions. While the impact of Western music was often seen as a potential musical death knell, it led to an unprecedented variety in music (Nettl 1985). The 20th century is thereby seen as the most intensive interchange of musical content worldwide, an unseen chance for composers, musicians, performers, (ethno)musicologists or any hybrid form of these.

References

Adler, G. 1885. Umfang, Methode und Ziel der Musikwissenshaft. *Vierteljahrsschrift Musikwissenschaft* 1: 5–20.

Agawu, K. 2006. Structural analysis or cultural analysis? Competing perspectives on the 'standard pattern' of West African rhythm. *Journal of the American Musicological Society* 59(1): 1–46.

Aubert, L. 2007. *The music of the other: new challenges for ethnomusicology in a global age.* Aldershot: Ashgate.

Cornelis, O., M. Lesaffre, D. Moelants and M. Leman. 2010. Access to ethnic music: advances and perspectives in content-based music information retrieval. *Signal Processing* 90(4): 1008–1031.

Dapper, O. 1668. *Naukeurige beschrijvingen der Afrikaensche gewesten*. Amsterdam: Jacob von Meurs.

Foster, H. 1995. The artist as ethnographer? In *The traffic in culture: refiguring art and anthropology*, ed. G.E. Marcus and F. Myers, 304–309. Berkeley: University of California Press.

Foucault, M. and P. Rabinow, eds. 1984. *The Foucault reader*. New York: Pantheon Books.

Koetting, J. 1970. Analysis and notation of West African drum ensemble. *Selected Reports in Ethnomusicology* 1(3): 115–147.

Krumhansl, C.L. 2010. Plink: 'thin slices of music'. *Music Perception* 27(5): 337–354.

Kunst, J. 1950. *Musicologica*. Amsterdam: Royal Tropical Institute.

Leman M. 2007. *Embodied music cognition and mediation technology.* London: The MIT Press.

Lesaffre, M., M. Leman, K. Tanghe, B. de Baets, H. de Meyer and J-P. Martens. 2003. User-dependent taxonomy of musical features as a conceptual framework for musical audio-mining technology. *Proceedings of the Stockholm Music Acoustics Conference*, Sweden, pp. 635–638.

Merriam A.P. 1977. Definitions of 'comparative musicology' and 'ethnomusicology': an historical-theoretical perspective. *Ethnomusicology* 21(2): 189–204.

Mersenne, M. 1636. *Harmonie universelle*. Paris: Cramoisy, imprimeur du Roy.

Moelants, D., O. Cornelis and M. Leman. 2009. Exploring African tone scales. *Proceedings of 10th ISMIR Conference*, Kobe, pp. 489–494.

Nettl, B. 1964. *Theory and method in ethnomusicology*. London and New York: The Free Press of Glencoe.

Nettl, B. 1985. *The impact on world music*. New York: Schirmer Books.

Oramas, S. and O. Cornelis. 2012. Past, present and future in ethnomusicology: the computational challenge. *Proceedings of the 13th ISMIR Conference*, Porto. http://ismir2012.ismir.net/event/papers/lbd11.pdf (accessed 8–12 October 2013).

Praetorius, M. 1619. *Syntagma Musicum III*. Wolfenbüttel: Elias Holwein.

Rousseau, J.J. 1768. *Dictionnaire de musique*. Amsterdam.

Sabbe, H. 1996. *All that music*. Leuven: Acco.

Sachs, C. 1955. *A short history of world music*. London: D Dobson.

Schwarz, D. 2006. Real-time corpus-based concatenative synthesis with CataRT. *Proceedings of the 9th Int. Conference on Digital Audio Effects (DAFx-06)*, Montreal, Canada, 18–20 September 2013.

Tzanetakis G., A. Kapur, W. Andrew Schloss and M. Wright 2007. Computational ethnomusicology. *Journal of Interdisciplinary Music Studies* 1(2): 1–24.

Wiora, W. and J. Gaudefroy-Demombynes. 1963. *Les quatre âges de la musique: de la préhistoire à l'ère de la technique*. Paris: Payot.

Lidy T., Silla C.N., Cornelis O., Gouyon F., Rauber A. ,Kaestner C.A.A., Koerich A.L. (2010). 'Western vs. Ethnic Music: On the Suitability of State-of-the-art Music Retrieval Methods for Analyzing, Structuring and Accessing Ethnic Music Collections', Signal Processing, Elsevier. 90 pp. 1032–1048.

Signal Processing 90 (2010) 1032–1048

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

On the suitability of state-of-the-art music information retrieval methods for analyzing, categorizing and accessing non-Western and ethnic music collections

Thomas Lidy [a,*], Carlos N. Silla Jr.[b], Olmo Cornelis [c], Fabien Gouyon [d], Andreas Rauber [a], Celso A.A. Kaestner [e], Alessandro L. Koerich [f]

[a] Vienna University of Technology, Austria
[b] University of Kent, Canterbury, UK
[c] University College Ghent, Belgium
[d] Institute for Systems and Computer Engineering of Porto, Portugal
[e] Federal University of Technology of Parana, Brazil
[f] Postgraduate Program in Informatics, Pontifical Catholic University of Paraná, Brazil

ARTICLE INFO

Article history:
Received 5 December 2008
Received in revised form
15 September 2009
Accepted 17 September 2009
Available online 23 September 2009

Keywords:
Ethnic music
Latin music
Non-Western music
Audio analysis
Music information retrieval
Classification
Access
Self-organizing map

ABSTRACT

With increasing amounts of music being available in digital form, research in music information retrieval has turned into a dominant field to support organization of and easy access to large collections of music. Yet, most research is focussed traditionally on Western music, mostly in the form of mastered studio recordings. This leaves the question whether current music information retrieval approaches can also be applied to collections of non-Western and in particular ethnic music with completely different characteristics and requirements.

In this work we analyze the performance of a range of automatic audio description algorithms on three music databases with distinct characteristics, specifically a Western music collection used previously in research benchmarks, a collection of Latin American music with roots in Latin American culture, but following Western tonality principles, as well as a collection of field recordings of ethnic African music. The study quantitatively shows the advantages and shortcomings of different feature representations extracted from music on the basis of classification tasks, and presents an approach to visualize, access and interact with ethnic music collections in a structured way.

1. Introduction

The availability of large volumes of music in digital form has spawned immense research efforts in the field of music information retrieval. A range of music analysis methods have been devised that are able to extract descriptive features from the audio signal. These are being used to structure large music collections, organize them into different categories, or to identify artists and instrumentation. They also serve as a basis for novel access and retrieval interfaces, allowing users to create personalized playlists, find preferred songs they would like to listen to or to interact with large music collections. Many of these approaches have by now found their way into commercial products.

However, most of this research has been carried out predominantly on Western music. This may be due to the

* Corresponding author.
E-mail addresses: lidy@ifs.tuwien.ac.at (T. Lidy), cns2@kent.ac.uk (C.N. Silla Jr.), olmo.cornelis@ugent.be (O. Cornelis), fgouyon@inescporto.pt (F. Gouyon), rauber@ifs.tuwien.ac.at (A. Rauber), celsokaestner@utfpr.edu.br (C.A. Kaestner), alekoe@ppgia.pucpr.br (A.L. Koerich).

0165-1684/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2009.09.014

easier availability of Western music in digital form. It may also reflect the larger familiarity of both the research community as well as the public in general with Western music, making evaluation of new approaches easier and leading to quicker industry take-up.

On the other hand, ethnic audio archives—collections of recordings from oral or tribal cultures—hold huge volumes of valuable music, collected by researchers all over the world over long periods of time. These form the basis of our musical cultural heritage. As a result of large and ongoing digitization and preservation projects, increasing volumes of ethnic music are becoming available in digital form, offering the basis for wider access and greater uptake. In order to fully unlock their value, these collections need to be made accessible with the same ease of use as current commercial music portals.

With ethnic music being in some aspects drastically different from Western music the question arises, in how far the research results stemming from traditional music information retrieval (Music IR) research can be directly applied. This question is not only posed for ethnic music collections but also for other non-Western music, such as Greek folk music, Latin American music, or traditional Indian music. Are the same audio description methods useful, although predominantly tested on music following Western tonality and rhythm principles? Does the optimization of these approaches to Western music benchmark collections lower applicability to non-Western music? Can comparable performance be obtained when trying to categorize ethnic and non-Western music automatically? Can Music IR provide tools and interfaces that allow researchers in ethnic music to access and evaluate their holdings in a sophisticated way, and may these interfaces also serve as an entry point for the general public, thus opening ethnic and cultural music collections to a larger user community?

Three music collections with different characteristics form the basis for detailed evaluations in this paper to address these questions. The first one is a common benchmark collection in Music IR research, consisting of predominantly Western style classical and Rock/Pop music, with some other genres such as World music mixed in. The second one is a collection of Latin American dance music, exhibiting dominating characteristics in terms of instrumentation and rhythm from a particular cultural domain, while still being strongly dominated by Western tonality and having been arranged and mastered using advanced studio technology. The third database consists of a collection of African ethnic music provided by the Ethnomusicological Sound Archive of the Belgian Royal Museum for Central Africa. This collection has drastically different recording standards, uses entirely different instruments and also has drastically different structures corresponding to music functions, geographic information, etc.

The tasks addressed in this article include specifically classification, where music is to be sorted automatically into various categories. These categories differ both in type (genre, instrumentation, geographical region, function) as well as their granularity. While classification of music is only one of many Music IR related tasks, it is also utilized for evaluation of the audio analysis methods that constitute the fundamental step of many other applications. We thus performed a systematic evaluation of a range of state-of-the-art audio feature extraction algorithms. Support vector machines (SVM) and ensemble classifiers based on time decomposition are used to evaluate performance differences in various settings. Apart from the automatic categorization of music archives we also present an interface to access music collections, based on self-organizing maps (SOM), that facilitates visual exploration and intuitive interaction with music collections and evaluated it regarding its suitability to help in the analysis and usage of ethnic music collections.

This article is organized as follows. Particular aspects to consider when working with ethnic music collections are described in Section 2. In Section 3, a review of the state-of-the-art in relevant fields of music information retrieval is given alongside previous related work on automatic analysis of ethnic and other non-Western music. Section 4 takes a detailed look on audio signal analysis and feature extraction methods that form the basis for the subsequent tasks. Section 5 then outlines the classification approaches used, describes the three characteristically distinct music databases used in the experiments in detail and presents comprehensive evaluation results on various classification strategies on the three databases. Section 6 presents the SOM-based access principles alongside a qualitative evaluation of music map interfaces based on the same three music collections. Conclusions are presented in Section 7, including remarks on issues to be addressed and an outlook on future work.

2. Peculiarities of ethnic music

Preparing an audio data set for Music IR-based research is not just about gathering available audio to build a collection. It needs a well-thought-out scheme of actions and intentions—considering both musical content and formal aspects—especially in the context of non-Western music.

While Western studio recordings are produced by specialists in an idealized environment with a clean song as a result and almost always no direct link between the producer and the consumer, ethnic music recordings are almost always made in the field and not in studio. They reflect a unique moment full of serendipity. Ethnic music is performed with a specific, often social, function serving its community, for instance with court songs, songs for rituals, songs for hunting, praise songs, work songs, etc. Western music is mainly produced for entertaining purposes. Behind the distribution of Western music generally a commercial motive is hidden [1], while for ethnic music it is passed through orally from generation to generation. Orally because there is no written culture, resulting in a musical framework that has neither defined rules nor concepts, an immense contrast with the Western music that relies on a very well-defined musical system. Because of these numerous differences, correct interpretation of ethnic music is not so evident, and researchers must always be aware not to pinpoint ethnic music on

Western musical theory or fall back on the existing musical concepts. Tzanetakis et al. [2] notice, for example, the opposition of Western music with its notion of a composition as a well-defined work to other music cultures where the boundaries between composition, variation, improvization and performance are more blurred and factors of cultural context have to be taken into account. Since ethnic music comes forward from an oral culture, a popular song can be brought by several, even dozens of local musicians, resulting in plural versions of one song, sometimes of varying quality, but often with personal influences affecting semantics, musical interpretation, instrumentation, and duration.

2.1. Musical content considerations

Audio from archival institutions can display diverging characteristics on pitch, temporal and timbral level if compared with commercial recordings. For instance, we can identify tendencies to differ in rhythmic aspects. Part of modern commercialized Western music (that is recorded, produced and mastered in studio) tends to show rhythmic aspects that are more controlled than in most ethnic music: deviations from perfect tempo and timing are most likely to be cautiously and systematically designed (or even avoided entirely), as opposed to ethnic music, more prone to emerging (bottom-up) tempo and timing deviations as well as to timing errors: for instance at the level of rubato or micro-timing [3,4], but also at a higher level such as the, probably intentional, constant speeding up during an entire song, or the, probably not intentional, slowing down of a group of singers if no percussion instruments are present.

Concerning meter, two main considerations have to be noticed: Western music handles a top-down approach starting with the major unit (a bar, or even a sentence), which is then rhythmically divided into smaller units, usually binary and ternary. Ethnic music tends to be organized additively: a bottom-up approach where the smallest unit can be seen as the starting point for extensions.

In African music timbre is an important aspect: since the variety of pitch and harmony is sober and the melodic lines are repetitive, other ways for enriching the sound are explored. For example, looking at African organology, the *lamallaephone* (also called thumb piano, sanza, ikembe) has some specific construction details by which its timbre range extends: every instrument is provided with a sound hole on the ventral side of the instrument. The performer can open and close this hole with his or her abdominal wall, generating a timbral and dynamic change of the sound. The lamellaephone is also often equipped with small metal rings that can vibrate when played. Another timbre-related aspect is the choice of material for building the lamellae, the soundboard and the sound box. A wide range of materials is being used to build musical instruments: specific materials can be wood, metal, turtle, reed, seeds, and in a few cases even human skulls have been used for the construction of the resonator, resulting in very different timbre, in spite of being the same instrument class. Concerning the representation of timbre related research, Western music has only few semantic labels dealing directly with timbre. Timbre is often referred to by descriptive terms, such as dark or brilliant, opaque or transparent, or in terms of metaphors, such as colors or moods. Frequently, Music IR applications avoid the semantic allocation of timbre by the use of similarity retrieval, recommendation or representations such as the SOM. For ethnic music that might also be the best way to handle timbre-related features.

The final remark is about the parameter pitch: when analyzed very precisely, the annotated frequencies that build the musical scale are deviating from the Western scales that are usually tuned according to the well-tempered 100 cents based 12 tone system. Representation of ethnic music scales are often referring to these Western note names, in the best case with their specific individual deviation mentioned. But it is conceptually wrong to try to relate the musical profiles onto the Western pitch classes [5].

2.2. Formal aspects

Ethnic music is usually the product of a field recording implying that its creation was in no case an optimal environment for achieving a perfect recording. The noise level can be very high, depending on the age of the recording, the recording and playback equipment, the deterioration of the original analogue carriers and the amount of time spent during the time-consuming digitalization process. Diverging levels of loudness can occur over separate collections or even within one collection. Some tracks even show unstable speed of the recording resulting in a pitch and tempo shift that is very

Table 1
African music database: functions and number of instances per function.

Festive song	48
Entertainment	178
Dance song	112
Work song	13
Narrative song	68
Evening song	8
Court music	45
War song	22
Religious song	14
Praise song	54
Historical song	25
Hunting	68
Cattle	44
Messages	21
Ritual music	79
Birth	19
Funeral	21
Lullaby	21
Mourning	26
Wedding	22
Narrative	4
Satire	1
Complaint	2
Riddle	1
Fable	1
Love song	4
Song of grace	4

hard to correct. Browsing an audio archive reveals the very diverging duration of audio tracks at first sight. While the shortest items in an archive are only a few seconds long, for example when presenting a scale or a very short message of a slit drum, the longest tracks overrun more than 1 h, dealing for example with a ceremony or a dance [6]. For some older collections, the beginning of audio tracks contains spoken information provided by the ethnomusicologist itself. A valuable attempt to provide audio of a physical attachment to its own meta-data, this unfortunately confuses common Music IR algorithms. An important remark concerning the meta-data of ethnic music compared to Western music is the availability and relevance of very different fields of information. Music from an oral culture will attribute no importance to composer, some importance to performer, but then again meta-data such as date and place of recording are more relevant. There is no genre label in a similar sense as the genres of Western music, rather the function of a song is an important attribute. Exemplary functions of an African ethnic music collection are given in Table 1, Section 5.2.3 contains a more specific description of the attributes of this collection.

3. Related work

In Western music, as opposed to what has been said about ethnic music in the previous section, the meta-data fields most frequently used (and searched for) are song title, name of artist, performer or band, composer, album, etc.—and a very popular additional one: the "genre" [7]. However, the concept of a genre is quite subjective in nature and there is no clear standard on how to assign a musical genre [8,9]. Nevertheless, its popularity has led to its usage not only in traditional music stores, but also in the digital world, where large music catalogues are currently labeled manually by genres. However, assigning (possibly multiple) genre labels by hand to thousands of songs is very time-consuming and, moreover, to a certain degree, dependent on the annotating person. Research in Music IR has therefore tackled this problem already in a variety of ways.

A brief analysis of the state of the art shows that there are different approaches in Music IR for the (semi-) automatic description of the content of music. In *content-based* approaches, the content of music files is analyzed and descriptive features are extracted from it. In case of audio files, representative features are extracted from the digital audio signal [10]. In case of symbolic data formats (e.g. MIDI or MusicXML), features are derived from notation-based representations [11]. Additionally, semantic *analyses of the lyrics* can help in the categorization of music pieces into categories that are not predominantly related to acoustic characteristics [12]. *Community meta-data* have also been used for such tasks, for instance, collaborative filtering [13], co-occurrence analysis (e.g. on blogs and other music related texts in the web [14,15]), or analysis of meta-information provided by users on dedicated third-party sources (e.g. social tags on last.fm [16]). In cases where manpower is available, *expert analyses* are an alternative and can provide powerful representations of music collections extremely useful for automatic categorizations (as in the case of Pandora[1] and the Music Genome Project,[2] or AMG Tapestry[3]). *Hybrid alternatives* also exist, they combine several of the previous approaches, e.g. combining audio and symbolic analyses [17], audio features, symbolic features and community meta-data [18] or combining audio content features and lyrics [19]. Although hybrid approaches have proved to be usually better than using a single approach, there are some implications on their use beyond traditional Western music. First of all, naturally, there is a lack of publicly available meta-data for non-Western and ethnic music, which could be used as a resource for hybrid approaches. Moreover, both community meta-data and lyrics-based approaches are dependent on natural language processing (NLP) tools, which are usually not in the same stage of development for English as opposed to other languages. Moreover, as seen in [20] the adaptation of an NLP method from one language to another is far from trivial. This is especially true for ethnic music where the NLP resources might not even exist.

While Music IR research has resulted into a wide range of methods and (also commercial) applications, non-Western music was rarely the scope of this research, and only little research has been performed with focus on ethnic music. Although ethnomusicology is a very traditional field of study with many institutions, both archival and academic, involved, research on the signal level has rarely been performed. Charles Seeger was one of the first researchers to objectively measure, analyze and transcribe sound, using his Melograph [21]. Later, pitch analysis on monophonic audio to score has also been performed by Nesbit et al. onto Aboriginal music [22]. Krishnaswamy focused on pitch contour enhancing annotations by assigning typologies of melodic atoms to musical motives from Carnatic music [23], a technique that is also employed by Chordia et al. on Indian music [24]. Moelants et al. point out the problems and opportunities of pitch analysis of ethnic music concerning the specific tuning systems differing from the Western well-tempered system [25]. Duggan et al. [26] analyzed pitch extraction results achieving segregation of several parts of Irish songs. Pikrakis et al. and Antonopoulos et al. performed meter annotation and tempo tracking on Greek music, and later also on African music [27,28]. Wright focuses on micro-timing of Rumba music, visualizing the smallest deviations of performance opposed to the transcription by the traditional theoretical musical framework [3]. A similar work on Samba music is done in [4]. Only very few authors presented work related to timbre and its usefulness in genre classification of ethnic music [29,30]. The term Computational Ethnomusicology was emphasized by Tzanetakis, capturing some historical, but mostly recent research that refers to the design, development and usage of computer tools within the context of ethnic music [2].

[1] http://www.pandora.com/
[2] http://en.wikipedia.org/wiki/Music_Genome_Project
[3] http://www.amgtapestry.com/

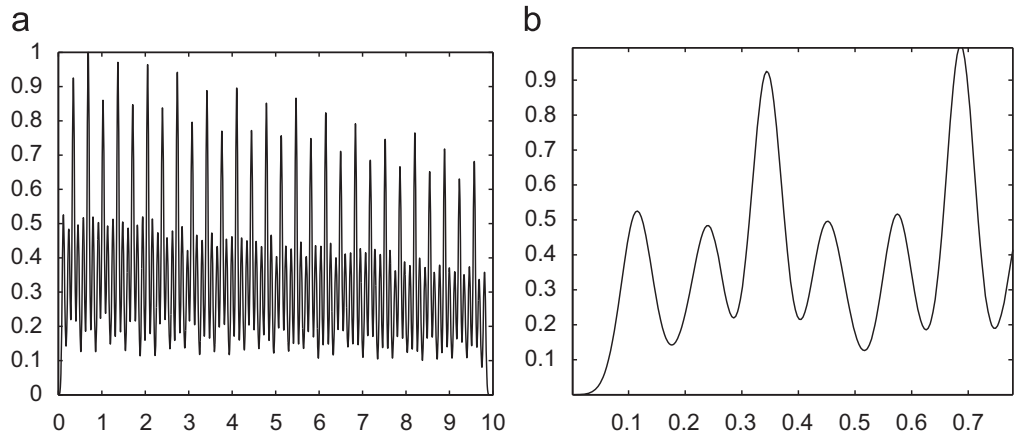


Fig. 1. Inter-onset interval histogram (IOIH): (a) IOIH [0,10] s, (b) IOIH detail view [0.0–0.78] s.

4. Audio analysis and feature extraction

A wealth of audio analysis and feature extraction methods has been devised in the Music IR research area for automated description of music [31]. Major approaches have been reviewed in Section 3 on related work. These feature extraction algorithms are employed in tasks such as automatic music classification, retrieval by (acoustic) similarity, or organization of music archives.

The set of algorithms used in our study comprises features from the MARSYAS framework developed by Tzanetakis et al. [10], inter-onset interval histogram coefficients (IOIHC) by Gouyon et al. [32], rhythm patterns (RP), by Rauber et al. [33] and its derivatives statistical spectrum descriptors (SSD) and rhythm histograms (RH) by Lidy et al. [34]. Additionally, two novel feature sets, based on SSD and RP features, are introduced in this article: temporal SSD and modulation variance descriptors (MVD). Following a brief description of all these feature extraction algorithms in Section 4.1 the creation of hybrid feature sets based on them is detailed in Section 4.2.

4.1. Feature extraction algorithms

4.1.1. MARSYAS features

The MARSYAS framework implements the original feature sets proposed by Tzanetakis and Cook [10]. The features can be divided into three groups: features describing the timbral texture (STFT and MFCC features), features for the rhythmic content (BEAT) and features related to pitch content (PITCH). The features for timbral texture are based on the short-time Fourier transform (STFT) and computed by the mean and variance of framewise spectral centroid, rolloff, flux, the time domain zero crossings, as well as the first five Mel-frequency cepstral coefficients (MFCCs) and low energy. Rhythm-related features aim at representing the regularity of the rhythm and the relative saliences and periods of diverse levels of the metrical hierarchy. They are based on a particular rhythm periodicity function: the so-called "beat histogram" (representing beat strength) and include statistics of the histogram (relative amplitudes, periods, ratios of salient peaks, as well as the overall sum of the histogram as an indication of beat strength). Pitch related features include the maximum periods of the pitch peak in the pitch histograms. The conjoint features form a 30-dimensional feature vector (STFT: 9, MFCC: 10, PITCH: 5, BEAT: 6).

4.1.2. Inter-onset interval histogram coefficients (IOIHC)

This pool of features tap into rhythmic properties of sound signals. They are computed from a particular rhythm periodicity function, the inter-onset interval histogram (IOIH) [35], that represents (normalized) salience with respect to period of inter-onset intervals present in the signal (in the range 0–10 s, cf. Fig. 1). The IOIH is further parameterized by the following steps: (1) projection of the IOIH period axis from linear scale to the Mel scale (by means of a filterbank), (2) IOIH magnitude logarithm computation, and (3) inverse Fourier transform, keeping the first 40 coefficients.

These steps intend to be an analogy of the Mel-frequency cepstral coefficients (MFCCs), but in the domain of rhythmic periodicities rather than signal frequencies. The resulting coefficients provide a compact representation of the IOIH envelope [32]. Roughly, lower coefficients represent the slowly varying trends of the envelope. It is our understanding that they encode aspects of the metrical hierarchy providing a high level view on the metrical richness, independently of the tempo. Higher coefficients, on the other hand, represent finer details of the IOIH. They provide a closer look at the periodic nature of this periodicity representation and are related to the pace of the piece at hand (its tempo, subdivisions and multiples), as well as to the rhythmical salience (i.e. whether the pulse is clearly established). This is reflected in the shape of the IOIH peaks: relatively high and thin peaks reflect a clear, stable pulse.

4.1.3. Rhythm pattern (RP)

A rhythm pattern is a set of features based on psycho-acoustical models, capturing fluctuations on frequency bands critical to the human auditory system [33,36]. In the first part, the spectrogram of audio segments of approximately 6 s (2^{18} samples) in length is computed using the short time fast Fourier transform (STFT) with a

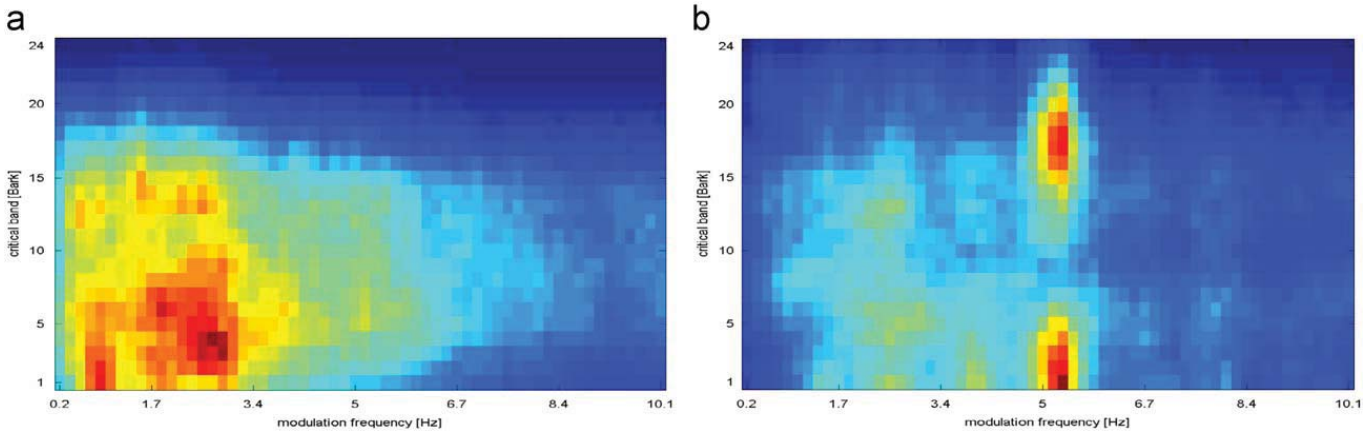


Fig. 2. Rhythm pattern: (a) Classical (b) Rock.

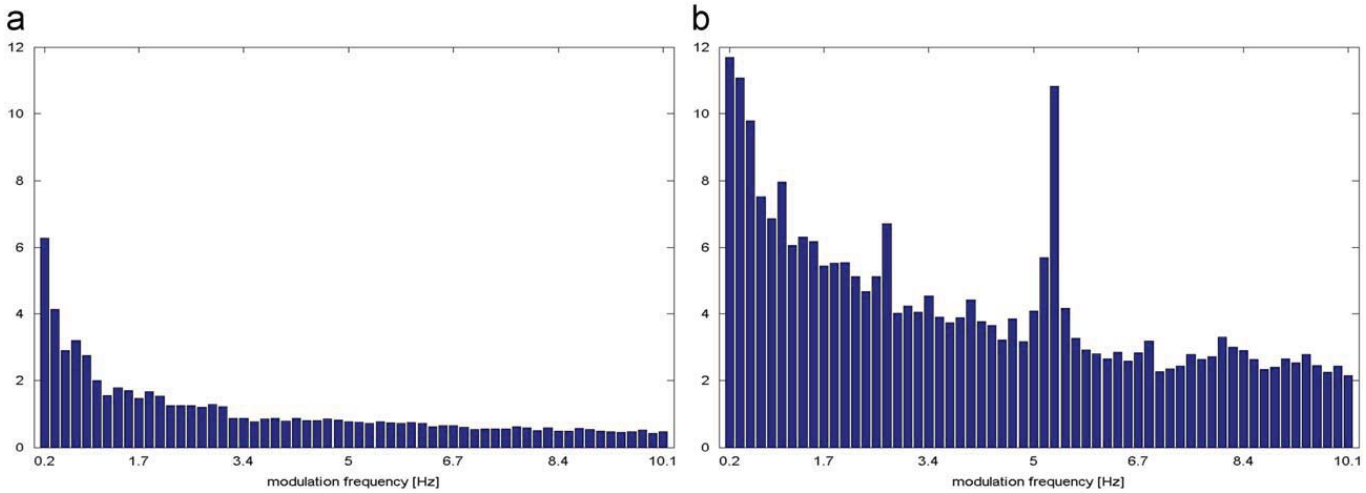


Fig. 3. Rhythm histograms: (a) Classical (b) Rock.

1024 samples[4] large Hanning window and 50% overlap. The Bark scale, a perceptual scale which groups frequencies to critical bands according to perceptive pitch regions, is applied to the spectrogram, aggregating it to 24 frequency bands [37]. The Bark-scale spectrogram is then transformed into the Decibel scale. Further psychoacoustic transformations are applied: computation of the Phon scale to incorporate equal loudness curves which account for the different perception of loudness at different frequencies and transformation into the Sone scale [37] to account for perceived relative loudness. The resulting Bark-scale Sonogram reflects the specific loudness sensation of an audio segment by the human ear.

In the second part, the varying energy on the critical bands of the Bark scale Sonogram is regarded as a modulation of the amplitude over time and its so-called "cepstrum" is retrieved by applying the Fourier transform. The result is a time-invariant signal that contains magnitudes of modulation per modulation frequency per critical band. This matrix represents a rhythm pattern, indicating occurrence of rhythm as vertical bars, but also describing smaller fluctuations on all frequency bands of the human auditory range. Subsequently, modulation amplitudes are weighted according to a function of human sensation of modulation frequency, accentuating values around 4 Hz, and cutting off frequencies >10 Hz. The application of a gradient filter and Gaussian smoothing improves similarity between rhythm patterns. The final 24×60 feature matrix is computed by the median of segmentwise rhythm patterns. Fig. 2 shows examples of a rhythm pattern for a classical piece and a rock piece. While the rock piece shows a prominent rhythm at a modulation frequency of 5.34 Hz, both in the lower critical bands (bass) as well as in higher regions (percussion, e-guitars), the classical piece does not exhibit such a distinctive rhythm but focuses on mid/low critical bands and low modulation frequencies.

4.1.4. *Rhythm histogram (RH)*

A rhythm histogram aggregates the modulation amplitude values of the 24 individual critical bands computed in a rhythm pattern (before weighting and smoothing), exhibiting the magnitude of modulation for 60 modulation frequencies between 0.17 and 10 Hz [34]. It is a lower-dimensional descriptor for general rhythmic

[4] For a sampling rate of 44,100 Hz; adjusted proportionally for lower rates.

characteristics in a piece of audio ($N = 60$, as compared to the 1440 dimensions of an RP). A rhythm histogram is computed for each 6 s segment in a piece of audio and the feature vector is then averaged by taking the median of the feature values of the individual segments (cf. Section 4.1.3).

Fig. 3 compares the rhythm histograms of a classical piece and a rock piece (the same example songs as for illustrating rhythm patterns have been used). The rock piece indicates a clear peak at a modulation frequency of 5.34 Hz while the classical piece generally contains less energy, having most of it at low modulation frequencies.

4.1.5. Statistical spectrum descriptor (SSD)

In the first part of the algorithm for computation of a statistical spectrum descriptor (SSD) the specific loudness sensation is computed on 24 Bark-scale bands (i.e. a Bark-scale Sonogram), analogously to a rhythm pattern. Subsequently, statistical measures are computed from each of these critical bands: mean, median, variance, skewness, kurtosis, min- and max-value, describing variations on each of the bands statistically. The SSD thus describes fluctuations on the critical bands and captures additional timbral information not covered by other feature sets, such as a rhythm pattern. At the lower dimension of 168 features this feature set is able to capture and describe acoustic content very well [34].

4.1.6. Temporal statistical spectrum descriptor (TSSD)

Feature sets are frequently computed on a per segment basis and do not incorporate time series aspects. As a consequence, TSSD features describe variations over time by including a temporal dimension. Statistical measures (mean, median, variance, skewness, kurtosis, min and max) are computed over the individual statistical spectrum descriptors extracted from segments at different time positions within a piece of audio. This captures timbral variations and changes over time in the audio spectrum, for all the critical Bark-bands. Thus, a change of rhythmic, instruments, voices, etc. over time is reflected by this feature set. The dimension is 7 times the dimension of an SSD (i.e. 1176).

4.1.7. Modulation frequency variance descriptor (MVD)

This descriptor measures variations over the critical frequency bands for a specific modulation frequency (derived from a rhythm pattern, cf. Section 4.1.3). Considering a rhythm pattern, i.e. a matrix representing the amplitudes of 60 modulation frequencies on 24 critical bands, an MVD vector is derived by computing statistical measures (mean, median, variance, skewness, kurtosis, min and max) for each modulation frequency over the 24 bands. A vector is computed for each of the 60 modulation frequencies. Then, an MVD descriptor for an audio file is computed by the mean of multiple MVDs from the audio file's segments, leading to a 420-dimensional vector.

4.2. Hybrid features

We make the hypothesis that a hybrid feature set combining multiple feature sets capturing, as much as possible, complementary characteristics of the music will achieve a better performance in retrieval and classification tasks.

A preliminary evaluation of the previously described individual feature sets on music databases with different characteristics showed also that some feature sets—to be more specific: certain feature attributes—are more discriminative on particular music collections than on others, depending on the musical content. This is a good incentive to try out diverse feature set combinations when dealing with Western vs. non-Western and ethnic audio collections.

Tzanetakis and Cook already proposed a hybrid feature set within the MARSYAS framework, i.e. the combination of STFT, MFCC, PITCH and BEAT features [10], called "MARSYAS-All" in this paper. These features represent multiple aspects of musical characteristics (namely, timbral, tonal and rhythmic). In this paper we propose to extend the hybrid approach by replacing the low-dimensional BEAT features in MARSYAS by the higher-dimensional ones described in Section 4.1, which are assumed to achieve more precise results because they capture a larger number of rhythmical and, for some of them, timbral aspects in the music. On the other hand, some of the feature sets have a strong focus on specific musical facets (e.g. rhythm) and might benefit vice versa from the conjoint feature sets. A number of hybrid feature sets is created, each based on Marsyas STFT, MFCC and PITCH + another feature set and the assumptions stated above are examined experimentally in Section 5.

5. Automatic music classification

A frequent scenario for the organization of audio archives is the categorization into a pre-defined list of categories (or, a related task, the assignment of class labels or tags). It is assumed that such a categorization, or classification, aids in managing an audio library. Based on audio feature extraction and a machine learning algorithm, classification of audio documents can be performed automatically. The machine learning research domain has developed a large range of classifier algorithms that can be employed. These algorithms are intended to find a separation of classes within the feature space. The approaches' premise is the availability of training data from which the learning algorithm induces a model to classify unseen audio documents.

In Section 5.1 we will briefly explain the classification approaches used in our study. Section 5.2 presents the data sets we used, containing Western, Latin American and African music. We will then present our experimental results on these databases in Section 5.3. We are investigating specific aspects of the Latin American and ethnic collections with regard to differences to classification of Western music, which is most frequently categorized into "genres".

5.1. Classification methods

5.1.1. Support vector machines

A support vector machine (SVM) [38] is a classifier that constructs an optimal separating hyperplane between two classes. The hyperplane is computed by solving a quadratic programming optimization problem, maximizing the distance of the hyperplane from its closest data vector. A "soft margin" allows a number of points to violate these boundaries. Except for linear SVMs the hyperplane is not constructed in the feature space but a kernel is used to project the feature vectors to a higher-dimensional space, in which the problem becomes linearly separable. Polynomial or radial basis function (RBF) kernels are common, however, for high-dimensional problems frequently a linear SVM performs equally or even better.

The sequential minimal optimization (SMO) algorithm is used in our approach, which breaks the large quadratic programming optimization problem of an SVM into a series of smallest possible problems, reducing both memory consumption and computation time, especially for linear SVMs [39].

5.1.2. Time decomposition

Combining multiple classifiers has been shown to improve efficiency and accuracy. Kittler et al. [40] distinguish from two different scenarios for classifier combination. In the first scenario, all the classifiers use the same representation of the input pattern. Although each classifier uses the same feature vector, each classifier will deal with it in different ways. In the second scenario, each classifier uses its own representation of the input pattern.

In this work, we employ the time decomposition (TD) approach [41,42], which is an ensemble-based approach tailored for the task of music classification that is related to the second scenario described above. TD can be seen as a meta-learning approach for the task of music classification as it is not dependent on any particular feature set or classifier. Feature vectors are frequently computed for individual segments of an audio document (cf. Section 4). When using this segmentation strategy it is possible to train a specific classifier for each one of the segments and to compute the final class decision from the ensemble of the results provided by each classifier. There are different ways to combine this information. In this paper we use majority voting (MAJ), the MAX rule (i.e. the output of the classifier with the highest confidence is chosen), the SUM and the PROD rules, where the probabilities for each class from each classifier are summed or multiplied, respectively, and the highest one is chosen.

5.2. Test collections

5.2.1. Western music database

As a reference we use a popular benchmark database of Western music. The collection was compiled for the genre classification task of the ISMIR 2004 Audio Description contest [43,44] and used frequently thereafter by Music IR researchers. The set of 1458 songs is categorized into six popular Western music genres: classical (640 pieces), electronic (229), jazz and blues (52), metal and punk (90), rock and pop (203), world (244). While the "world" music genre partially covers non-Western music as well, this coarse genre subdivision is typical for average users of Western music collections.

5.2.2. Latin music database

The Latin music database (LMD) [45] contains 3227 songs, which were manually labeled by two human experts who have over 10 years of experience in teaching Latin American dances. The data set is categorized into 10 Latin music genres (Axé, Bachata, Bolero, Forró, Gaúcha, Merengue, Pagode, Salsa, Sertaneja, Tango). Contrary to popular Western music genres, each of these genres has a very specific cultural background and is associated with a different region and/or ethnic and/or social group. Nevertheless, it is important to note that in some aspects the Latin music database is musically similar to the Western music database as it makes use of modern recording and post-processing techniques. By contrast to the Western music database, the LMD contains at least 300 songs per music genre, which allows for balanced experiments.

5.2.3. African music database

The collection of African music used in this study is a subset of 1024 instances of the audio archive of the Royal Museum of Central-Africa (RMCA)[5] in Belgium, kindly provided by the museum. It is one of the largest museums in the world for the region of Central-Africa, with an audio archive that holds 50,000 sound recordings from the early 20th century until now. This unique collection of cultural heritage is being digitized in the course of the DEKKMMA project [46], one goal being to provide enhanced access through the use of Music IR methods [47]. There is a lot of meta-data available for the collection, related to identification (number/id, original carrier, reproduction right, collector, date of recording, duration), geographic information (country, province, region, people, language), and musical content (function, participants, instrumentation). Unfortunately, not for every recording all fields are available, as often these data cannot be traced.

A number of these meta-data can be used to investigate the methods of Music IR for classification and access. One important meta-data field investigated in this study is the "function", describing specific purposes for individual pieces of music. Table 1 shows the number of instances for the 27 different functions available in the collection. The database is partially annotated by instrumentation, with a 3-level hierarchy and a single or multiple instruments per song. Level 1 is a categorization by instrument family, on the second level there were 28 different instruments in the database, with an optional subtype on the third level. Instrument families and instruments on level 2 are given in Table 2.

Another category investigated was the country. The list of countries can be seen in Table 9. The database contains also a field with the name of the people (ethnic group) who played the music. 693 instances have been annotated

[5] http://music.africamuseum.be

Table 2
African music database: instrument families and instruments.

Aerophone	Flute, flute (European), horn, pan pipe, whistle, whistling
Chordophone	Fiddle, guitar, harp, lute, musical bow, zither
Idiophone	Bell, handclapping, lamellaphone, percussion pot, pestle, rattle, rhythm stick, scraper, sistrum, slit-drum, struck object, xylophone
Membranophone	Drum, friction drum, single-skin drum, double-skin drum

with an ethnic group, in total 40 different ethnic groups are known in the database.

5.3. Experimental results

We investigated the classification of audio documents measuring accuracy on a multi-class classification task. The Weka Machine Learning tool [48] was employed in all experiments, using the SMO algorithm, and, in a subset of experiments, the time decomposition approach on top of it. Linear SVMs were trained, with the complexity parameter c set to 1. All experiments were run using stratified 10-fold cross-validation. Potential improvements of the time decomposition ensemble approach over a single SVM were investigated, as well as a comparison of analyzing different audio segments. Apart from the latter experiment, all experiments are based on a feature analysis of the center 30 s of the pieces in the music collections. Results are presented as accuracy values in percent and the standard deviation. Though the numerical results cannot be directly compared between the three databases, due to different organization schemes and semantics (e.g. genre vs. function) as well as different sizes of the collections and different numbers of classes, these results allow an assessment of how well the approaches are also applicable to non-Western music archives.

5.3.1. Results on the Western music database

Feature set comparison: The first experiment includes a comparison of the dependence of the results on the particular segment taken as excerpt for analysis from the audio signal. Three different 30-s audio segments were analyzed: the beginning, the center and the end part of each piece ($Seg_{beg}, Seg_{mid}, Seg_{end}$). Table 3 shows the results of this segmentwise analysis for all the different feature sets described in Section 4. The results indicate a rather moderate performance of rhythm and beat related features (e.g. RH, MARSYAS-BEAT, IOIHC) while other feature sets that capture more timbral information achieve higher results, with a classification accuracy of 76.12% using SSD features.

The MARSYAS-BEAT features have been replaced successively by other feature sets. The lower part of Table 3 presents the results for these hybrid feature sets, where the combination of MARSYAS (excl. BEAT) features with SSD achieved 79% accuracy on Seg_{mid}. In general, the hybrid approaches performed always better than the individual approaches, with all results based on the middle audio segment being higher than 71%. The MARSYAS-All combination was improved in all but two cases (HYBRID-IOIHC, HYBRID-RH). The results imply that the additional feature sets capture musical (both rhythmic an timbral) aspects better than the MARSYAS-BEAT features.

Table 3
Western music database: segment comparison (SVM classification by genre).

Feature set	Seg_{beg}	Seg_{mid}	Seg_{end}
MARSYAS-STFT	56.36±1.42	61.72±2.28	59.54±1.84
MARSYAS-PITCH	45.66±1.89	52.49±2.42	49.70±2.22
MARSYAS-MFCC	58.19±2.34	65.47±2.47	60.90±2.80
MARSYAS-BEAT	52.06±2.34	54.87±2.15	53.55±2.57
IOIHC	45.00±1.56	49.71±1.31	42.61±1.10
RH	57.55±1.86	62.84±2.53	59.11±2.86
RP	64.94±3.95	69.78±3.30	64.81±3.91
SSD	**71.20**±2.52	**76.12**±3.76	**72.71**±2.42
TSSD	64.02±4.20	70.14±3.39	65.76±2.28
MVD	62.88±2.51	68.47±1.75	62.73±2.40
MARSYAS-All	66.57±3.03	71.85±2.62	67.54±2.29
HYBRID-IOIHC	64.08±3.14	71.50±2.07	64.48±3.20
HYBRID-RH	66.87±2.63	71.69±2.12	68.94±1.94
HYBRID-RP	71.03±4.47	75.13±2.98	71.65±2.53
HYBRID-SSD	**75.40**±3.03	**79.00**±3.13	**76.07**±2.94
HYBRID-TSSD	68.64±4.39	74.85±3.25	69.74±2.30
HYBRID-MVD	68.76±2.11	73.26±2.34	68.02±1.27

Segment comparison: The highest accuracy for Seg_{beg} is 75.40% with HYBRID-SSD features. For the Seg_{end} the highest accuracy is 76.07% with HYBRID-SSD features. The comparison among Seg_{beg}, Seg_{mid} and Seg_{end} shows that extraction of features from the center segment (Seg_{mid}) performs better in all cases. This comparison was performed also for the Latin music collection with quasi the same outcome, we therefore omit presenting the segment-analysis results for the Latin music database in the next section.

By contrast to the rather clear conclusion on the Western and Latin music databases, where there is a difference of up to 5 percentage points using Seg_{beg} or Seg_{end} instead of Seg_{mid}, the situation is different with the African music database, as will be shown in Section 5.3.3.

Time decomposition: The results presented in Table 4 are based on the ensemble of the features extracted from Seg_{beg}, Seg_{mid} and Seg_{end}. The overall highest result on individual feature sets was 77.20% using SSD features and the majority vote (MAJ) rule (the result using a single SVM was 76.12%). Overall, time decomposition (TD) improved the results for five of the feature sets, but in three cases the results were marginally worse than using linear SVM only. However, the best results were generated using different ensemble voting rules, with the MAX rule being most frequently the best rule, although SUM and PROD seem to be better when using hybrid feature sets.

The highest overall result is 80.37% using HYBRID-SSD features and the majority vote rule. The TD approach generally improved results for hybrid features only very moderately, except for H-RP and H-TSSD features, where the SUM and PROD rules achieved improvements by about 3.4 percentage points.

Table 4
Western music database: classification using the time decomposition approach.

Feature set	Ensemble rule			
	MAJ	MAX	SUM	PROD
MARSYAS-STFT	60.91±1.27	61.58±2.18	60.78±1.33	61.21±1.65
MARSYAS-PITCH	50.34±1.45	52.49±2.30	49.45±1.40	49.45±1.36
MARSYAS-MFCC	63.26±2.50	65.12±2.59	63.92±2.59	63.84±2.83
MARSYAS-BEAT	54.73±2.64	54.65±2.02	54.60±3.34	53.86±1.72
IOIHC	45.00±1.60	49.71±1.31	45.07±1.52	45.07±1.52
RH	60.81±2.06	62.84±2.44	61.58±2.51	61.18±2.11
RP	70.99±3.23	70.39±3.55	72.40±2.61	72.00±2.31
SSD	**77.20**±3.28	**76.19**±3.80	**76.52**±2.57	**76.73**±2.62
TSSD	72.39±2.38	70.34±3.58	73.42±2.17	72.86±2.99
MVD	67.63±2.01	69.01±2.29	68.65±2.69	68.52±2.72
MARSYAS-All	71.44±2.45	72.21±2.59	71.42±2.26	71.14±1.69
HYBRID-IOIHC	69.51±2.83	71.64±1.96	69.66±1.47	69.65±1.19
HYBRID-RH	70.80±2.73	71.90±2.07	71.45±2.31	72.38±1.93
HYBRID-RP	77.12±3.49	75.54±3.40	78.34±2.54	78.57±2.46
HYBRID-SSD	**80.37**±2.65	**79.02**±2.81	**79.75**±2.65	**79.54**±2.61
HYBRID-TSSD	78.18±2.22	75.23±3.24	78.25±3.35	77.74±3.46
HYBRID-MVD	73.49±1.05	73.60±1.98	73.80±2.32	73.65±2.72

Table 5
Latin music database: comparison of SVM and the time decomposition approach.

Feature set	SVM	Time decomposition	
	Seg_{mid}	MAJ	SUM
MARSYAS-STFT	56.40±2.13	57.73±2.58	56.93±2.32
MARSYAS-PITCH	25.83±1.63	27.33±0.96	29.53±2.17
MARSYAS-MFCC	58.83±2.31	60.20±2.02	60.26±2.86
MARSYAS-BEAT	31.86±1.69	33.40±1.48	34.56±2.43
IOIHC	53.26±2.63	52.53±2.74	47.73±2.76
RH	54.63±1.94	56.96±2.03	57.80±2.27
RP	81.40±1.45	**84.76**±1.23	**84.70**±1.25
SSD	**82.33**±1.36	84.70±1.50	84.06±1.33
TSSD	73.80±1.75	79.40±2.23	81.70±1.11
MVD	67.70±2.75	71.66±2.03	73.00±1.96
MARSYAS-All	68.46±2.03	70.40±2.23	70.40±1.99
HYBRID-IOIHC	77.63±1.74	78.33±1.82	77.13±1.67
HYBRID-RH	74.50±2.47	76.73±2.19	77.16±2.19
HYBRID-RP	84.06±1.42	87.46±1.66	**88.06**±1.60
HYBRID-SSD	**85.30**±1.39	**87.53**±1.20	87.40±1.01
HYBRID-TSSD	75.80±1.93	81.93±2.33	83.96±1.58
HYBRID-MVD	77.06±2.71	81.50±1.56	81.50±1.19

5.3.2. Results on the Latin music database

Feature set comparison: Classification with a single SVM compared by using different audio segments was always better using the center 30 s of a song (Seg_{mid}). Therefore, results on the beginning and end segments are omitted in Table 5. It seems that both rhythm and timbre play a major role in discriminating Latin music genres, with rhythm patterns (RP) and SSD giving the best results (81.4% and 82.33%, respectively). It is especially noteworthy that pitch seems to play a subordinate role noticeable by the low performance of MARSYAS-PITCH features (25.83%). Pure rhythmic features deliver intermediate results (BEAT, IOIHC, RH). The hybrid approaches bring a major boost to them. This is explainable having a look at the Latin American genres in the database where some genres have a similar rhythm. For example, Forró, Pagode, Sertaneja and Gaúcha are rhythmically similar and for that reason the other features help to distinguish between them. There are also similarities between Bolero and Tango. Additional evidence for these similarities is presented in the confusion matrix in Table 6 where a large portion of Bolero is misclassified as Tango, Sertaneja is confused with Bolero, numerous Pagode songs are misclassified as Salsa, Bolero, Gaúcha or Sertaneja, and many Forró songs are confused with Gaúcha.

Table 6
Confusion matrix for Latin music genres, using RH features and SVM.

	Ta	Sa	Fo	Ax	Ba	Bo	Me	Ga	Se	Pa
Tango	263	0	2	0	0	26	0	7	1	1
Salsa	15	158	7	17	4	22	3	29	19	26
Forró	16	15	130	19	1	6	9	59	28	17
Axé	15	34	45	89	4	6	49	35	14	9
Bachata	5	1	1	4	237	4	28	7	9	4
Bolero	66	5	0	1	2	174	1	3	38	10
Merengue	1	2	13	33	17	2	211	8	13	0
Gaúcha	26	15	34	26	9	21	7	129	18	15
Sertaneja	6	16	29	18	28	59	9	15	104	16
Pagode	15	30	13	18	0	31	2	29	30	132

The hybrid sets are significantly better than the MARSYAS-All approach in all cases. The addition of more complex features to the MARSYAS set instead of the BEAT features achieved a major increase in classification accuracy. The major trends are similar to the Western music database, although the specific combination of rhythmic and timbral characteristics in the RP features seems to be of particular use for Latin American music.

Time decomposition: The ensemble approach could increase the performance by several percent. The MAX and PROD rules were generally inferior to the MAJ and SUM rules and are therefore not presented. An interesting

aspect is that RP features surpassed SSD using time decomposition, a hint to more individual rhythmic characteristics extracted from individual segments. The same effect appears with HYBRID-RP features on the SUM rule, where the accuracy is as high as 88.06% (a 4% improvement over the single SVM).

5.3.3. Results on the African music database

The many meta-data fields available for the African music database allow classification by multiple facets. The results give an assessment about what kind of information can be detected by current feature analysis and classification approaches and potential challenges specific for classification of ethnic music.

Segment comparison: Before performing classification by different categories, we have carried out a pre-analysis of the performance of different audio segments, as we have done for the Western and Latin music databases. We have carried out classification with SVM considering the function as classes. From Table 7 it is visible that, in general, there is much less difference between the different 30-s segments used for analysis, with deviations below 2 percentage points; for some feature sets, the performance is even equal. However, some other feature sets seem to perform particularly well on the initial segment (Seg_{beg}). This might be a hint that the beginning of ethnic music pieces may be important for characterizing its function, a circumstance that is the contrary with Western music and its frequent lead-in effects. The end segment provides mainly worse results than the center segment, which is similar to Western and Latin music, yet not to the same extent, pointing at a lower presence of fade-out effects in ethnic music. Generally, there is yet again the conclusion that usually the inner content of a piece of music contains more characteristics useful for classification. In the subsequent classification experiments, Seg_{mid} is used for evaluation.

Table 7
African music database: segment comparison (SVM classification by function).

Feature set	Seg_{beg}	Seg_{mid}	Seg_{end}
MARSYAS-STFT	21.86±2.63	23.14±2.88	21.92±2.50
MARSYAS-PITCH	21.09±2.22	21.09±2.22	21.09±2.22
MARSYAS-MFCC	30.22±3.91	27.19±4.44	26.36±3.53
MARSYAS-BEAT	21.09±2.22	21.22±2.30	21.09±2.30
IOIHC	21.07±2.28	21.07±2.28	20.81±2.33
RH	21.39±3.07	22.10±2.49	21.84±2.77
RP	36.83±3.42	37.85±5.49	35.29±4.03
SSD	**44.27**±6.63	**45.12**±5.98	**44.34**±4.34
TSSD	37.16±5.23	35.75±3.23	37.14±6.43
MVD	27.76±3.12	28.28±5.66	28.48±4.50
MARSYAS-All	35.38±4.57	32.39±5.56	30.18±4.47
HYBRID-IOIHC	36.72±4.47	30.34±5.45	29.90±5.03
HYBRID-RH	34.21±8.06	34.93±5.33	33.83±4.96
HYBRID-RP	39.63±5.88	41.12±5.81	38.82±3.02
HYBRID-SSD	46.46±6.33	48.25±6.63	47.24±4.27
HYBRID-TSSD	38.81±4.93	38.10±2.66	39.37±7.03
HYBRID-MVD	33.69±7.37	35.08±4.63	35.65±6.84

Classification by function: The "function" field in the African music database was of specific interest to us, because it may be considered as the counterpart of the genre in Western music. From the originally 27 different functions available in the set (cf. Table 1) functions with less than 10 instances were ignored for the following experiments, in order to permit a proper cross-validation, keeping 19 functions. The second column of Table 8 provides the results of classification by function using the different feature sets and hybrid approaches (all based on Seg_{mid}). The accuracies achieved were rather low (the baseline considering the largest class is 19.7%); it seems that the concept of a "function" is not captured very well by the audio analysis methods used. The only prominent results are delivered by the SSD features, with 45.12% accuracy, followed by RP features with 37.85% and temporal SSD with 35.75%.

The use of the hybrid feature sets shows an improvement in accuracy well over the baseline for all feature sets compared to the original individual feature sets. They also show an improvement over the MARSYAS-All combination

Table 8
African music database: classification by different meta-data using SVM.

Feature set	Function	Instrument	Country	Ethnic group
MARSYAS-STFT	23.14±2.88	38.71±4.77	58.61±6.62	61.59±9.34
MARSYAS-PITCH	21.09±2.22	36.85±3.92	47.20±3.85	60.53±9.24
MARSYAS-MFCC	27.19±4.44	46.87±8.31	54.64±4.44	70.96±10.84
MARSYAS-BEAT	21.22±2.30	37.64±4.46	50.25±6.16	60.53±9.24
IOIHC	21.07±2.28	39.14±6.88	55.19±6.95	60.53±9.24
RH	22.10±2.49	44.34±5.23	62.17±6.73	63.41±9.70
RP	37.85±5.49	57.42±5.48	72.24±5.31	80.57±5.31
SSD	**45.12**±5.98	67.61±7.87	**81.74**±4.70	**85.07**±5.07
TSSD	35.75±3.23	**69.06**±6.66	81.21±3.76	84.88±5.12
MVD	28.28±5.66	47.90±7.83	65.22±3.63	68.41±8.08
MARSYAS-All	32.39±5.56	55.44±7.72	64.24±5.00	76.62±6.93
HYBRID-IOIHC	30.34±5.45	59.85±7.04	68.59±5.82	79.96±7.08
HYBRID-RH	34.93±5.33	59.75±10.23	72.61±5.02	81.01±7.14
HYBRID-RP	41.12±5.81	64.11±5.37	77.00±5.25	84.88±5.43
HYBRID-SSD	**48.25**±6.63	68.79±7.77	**82.21**±3.34	88.10±3.75
HYBRID-TSSD	38.10±2.66	**68.82**±4.88	80.71±4.16	**88.57**±3.89
HYBRID-MVD	35.08±4.63	57.59±8.52	75.03±3.41	77.55±6.29

in all except the HYBRID-IOIHC case. However, the improvement of the best result (SSD) was moderate, with HYBRID-SSD achieving 48.25%. Evaluation of the time decomposition approach showed that the highest result on HYBRID-SSD could be further improved to 50.05% using the SUM rule.

Classification by instrumentation: We have conducted an experiment on classification of the instrument family. 711 instances were labeled by instrumentation. A mixture of multiple instruments per song was considered as a separate class (see class list in Fig. 7). Instrument recognition naturally is not a task done with rhythmic features as can be also seen from column 3 in Table 8. Better results are accomplished clearly by the timbral features, with MFCC achieving 46.87%, SSD 67.61% and temporal SSD 69.06%. The hybrid set-up could improve the performance of the low-performing rhythm-based feature sets, but not the one of TSSD features and only slightly the one of SSD.

Classification by country: Our subset of the collection contained music from 11 African countries. The countries Eritrea and Niger were, however, represented by one piece each only and were thus ignored for this experiment as it would be impossible to create a training and test set for these.

The results of the classification by country (column 4 of Table 8) indicate that the audio features under investigation allow a proper classification into the originating countries of the African audio recordings. The results indicate that the rhythmic properties differ between the countries (with RH and RP features giving quite high results) but also that timbral aspects play a major role, probably due to the use of different instruments: SSD achieved an accuracy of 81.74% and TSSD 81.21%. Hybrid feature sets could improve these results only marginally.

The confusion matrix in Table 9 shows that the major confusion happens only between Congo DRC and Rwanda. Being geographical neighbors, these countries' cultures are very related and so is their music. Even with the dominance of audio instances available from the Congo DRC and Rwanda, classification of pieces of audio from the other less represented countries performed very well, with 100% recall and precision on classifying music from the Ivory Coast. Overall, precision and recall were 72.8% and 67.5%, respectively.

Table 9
Confusion matrix for African music by country, using TSSD features and SVM.

	Rw	Bu	Co	Ga	RC	Et	Se	Gh	Iv
Rwanda	324	5	65	1	2	0	1	0	0
Burundi	11	6	0	0	0	0	0	0	0
Congo DRC	74	0	384	0	0	0	0	2	0
Gabon	1	0	0	6	4	0	0	0	0
Republic of the Congo	1	0	0	4	7	1	1	0	0
Ethiopia	4	0	5	1	1	12	2	0	0
Senegal	1	0	0	0	1	1	7	0	0
Ghana	2	0	3	0	0	0	0	27	0
Ivory Coast	0	0	0	0	0	0	0	0	15

Classification by ethnic group: In this experiment, we have investigated whether our classification approach is able to separate the music according to the ethnic group that performed it. The baseline for this experiment was 50.22% as there were 348 instances from the "Luba" people. Column 5 of Table 8 shows that the feature sets based mainly on rhythm could not distinguish the music very well by ethnic group, but feature sets incorporating timbral aspects achieved remarkable results on classifying 40 ethnic groups: RP achieved 80.57% classification accuracy, SSD features accomplished remarkable 85.07%, and the TSSD features reached 84.88%. The hybrid approaches could further improve these results to an astounding classification accuracy of more than 88%.

Impressed by these high results, we tried to investigate whether there may be a direct correlation between the recording quality of the pieces of a specific ethnic group. Unfortunately, there was no reference to the recording equipment that was used at the time of the recording of the pieces in the database. From the meta-data we had available we could, however, investigate (1) whether there is an influence introduced by the bitrate used for encoding the recordings and (2) whether there is any correlation between the year of the recording and the ethnic group. Over all the different recordings, only three different MP3 bitrates were used: 128, 192 and 256 kbit/s, so the effect of the bitrate should be negligible. The recordings for 40 different ethnic groups were made in 13 different years between 1954 and 1975, and some in 2005 and 2006. From listening to the recordings we could not perceive major quality differences between the recordings. More importantly, the recordings of a specific ethnic group were not made in a single year, e.g. the music of the Twa people was recorded in 1954, 1971 and 1973–1975. It is plausible that not the same equipment was used in all these years. On the other hand, in several individual years, multiple ethnic groups have been recorded, potentially with the same equipment. Thus, there does not seem to be evidence for any correlation between recognition of ethnic group and recording equipment.

On the other hand, in general, it is hardly possible to avoid that potential recording effects influence the classification results. However, exactly the same is true for Western music, where the instrumentation, voice, etc. of a specific performer and/or the mastering of a certain producer can have an effect on the classification results.

6. Alternative methods of access to music collections

Classification into pre-defined categories faces a particular issue: the definition of the categories. Although classification seems like an objective task, the definition of categories, no matter if done by experts or users of a private music collection, is subjective in its nature. As a consequence, the defined categories are overlapping—a fact that can frequently be observed particularly with Western musical genres—and there are no clear boundaries between the categories, neither for humans, nor for a machine classifier.

This problem is especially prevalent in collections of ethnic audio documents where the concept of a "genre" is

frequently inexistent. The African music database described in Section 5.2.3, for example, contains a "function" category which describes a situation where a song is played rather than a genre in the sense of Western music. Especially for an automatic classification system it is therefore more difficult to determine the "function" of a song by acoustic content than a genre, which is supposed to be, to a certain extent, distinctive by sound. Commonly, a function is also related to the lyrics of a song.

With the lack of the concept of a genre defined by similar musical and sound characteristics, the question arises of how to structure and access ethnic music collections. When the ethnic music collection is thoroughly labeled and entered into a database system it is possible to retrieve music by searching or ordering the available meta-data fields. However, even with meta-data such as function, country or people retrieval by acoustically similar groups is difficult.

Using the concept of self-organizing maps, which organize music automatically according to acoustic content, access by acoustic similarity can be provided to ethnic music which would otherwise not be possible. In the following sections we will describe the underlying principles and a software application that provides this kind of access to music collections by sound similarity.

6.1. The self-organizing map

There are numerous clustering algorithms that can be employed to organize an audio collection by sound similarity based on different acoustic features. An approach that is particularly suitable is the self-organizing map (SOM), an unsupervised neural network that provides a mapping from a high-dimensional input (feature) space to a (usually) two-dimensional output space [49].

A SOM is initialized with an appropriate number of units, arranged on a two-dimensional grid. A weight vector is attached to each unit. The input space is formed by feature vectors extracted from the music. The vectors from the (high-dimensional) input space are randomly presented to the SOM and the activation of each unit for the input vector is calculated using, e.g. the Euclidean distance between the weight vector of the unit and the input vector. Next, the weight vector of the activated unit is adapted towards the input vector. Consequently, the next time the same input signal is presented, the unit's activation will be even higher. The weight vectors of neighboring units are also modified accordingly, yet to a smaller amount. The magnitude of modification of the weight vectors is controlled by a time-decreasing learning rate and a neighborhood function.

This process is repeated for a large number of iterations, presenting each input vector multiple times to the SOM. The result of this training procedure is a topologically ordered mapping of the presented input data in the two-dimensional space. Similarities present in the input data are reflected as faithfully as possible on the map. Hence, similar sounding music is located close to each other, building clusters, while pieces with more distinct acoustic content are located farther away. Clearly distinguishable musical styles (e.g. distinctive genres) will be reflected by cluster boundaries, otherwise the map will reflect smooth transitions among the variety of different pieces of music.

6.2. The application

Based on the SOMeJB system [36] that extended the purely analytical SOM algorithm by advanced visualizations, the PlaySOM application enhances the principle to a rich application platform that provides direct access to the underlying music database enriched by browsing, interaction and retrieval [50]. The application's main interface visualizes the self-organized music map in one of many different visualization metaphors (acoustic attributes, "Weather Charts", "Islands of Music", among various others). It provides a semantic zooming facility which displays different information dependent on the zooming level. The outer zooming level provides a complete overview of the music collection, with numbers indicating the quantity of audio documents mapped at each location. The default visualization, smoothed-data-histograms [51], indicate clusters of music that have coherent acoustic features. Zooming into the map, more information is shown about the individual audio titles.

A search window allows querying the map for specific titles. The benefit of organizing the music collection on a SOM is that similar sounding pieces of music can be retrieved directly by exploring the surroundings of the unit where a searched item has been retrieved from. With the same ease of clicking into the map, a playlist is created on-the-fly. Marking a rectangle selects an entire "cluster" of music that is perceived as acoustically similar, or a subset of the audio-collection matching a particular musical style. A path selection mode allows drawing trajectories through the musical "landscape" and selects all pieces belonging to units beneath that trajectory. This allows creating ad hoc music playlists with (smooth) transitions between various musical styles. These immediate selection and playback modes are particularly useful for a quick evaluation of the clustered content of such a music map. Variants of the PlaySOM application have been created for a range of mobile devices [52], a platform, which is in particular need for enhanced access methods not based on traditional genre-album-artist lists.

6.3. Experimental results

Although the SOM principle and the PlaySOM application are not based on external meta-data at all, an overlay visualization of genre or class meta-information on top of the music map is available. This form of visualization helps in analyzing the experimental results by showing class assignments as pie-charts on top of each SOM unit, using different colors to depict different classes. This will thus provide an implicit kind of evaluation of the automatic organization of a music map by acoustic similarity: the more coherent the colors are on top of

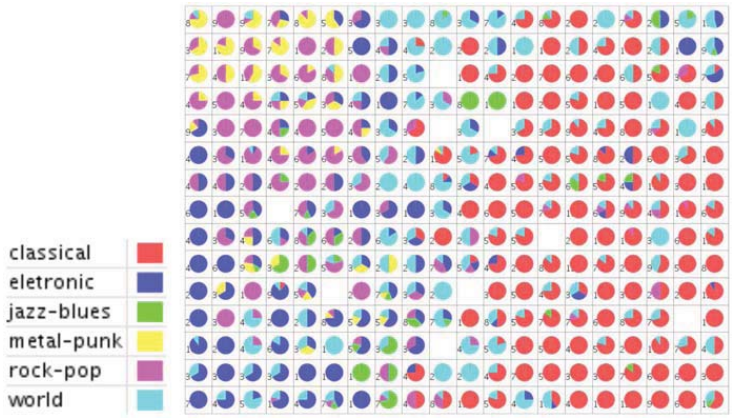


Fig. 4. Map of Western music database, visualized by genre.

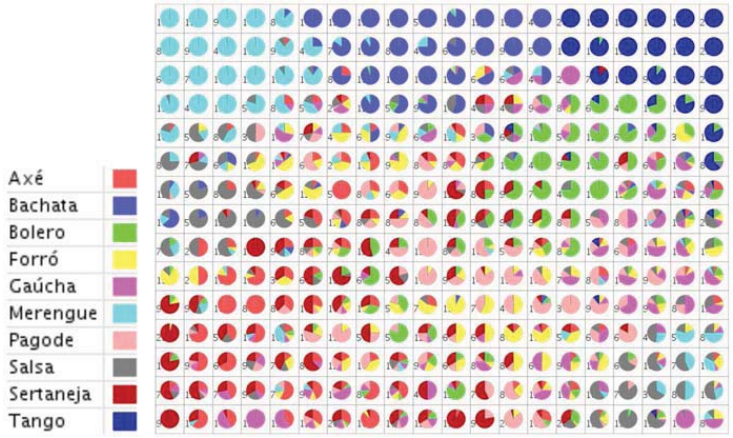


Fig. 5. Map of Latin music database, visualized by genre.

the map, the more it agrees with manual human annotation.[6]

For each of the three music collections studied already in Section 5, a 20×15 SOM was created using SSD features (cf. Section 4.1.5) extracted from the audio as input. Fig. 4 shows the Western music collection automatically aligned by audio content on the units of a SOM. Each unit shows a pie-chart class diagram indicating in various colors the portions of each of the six classes listed in Fig. 4. The number of songs per unit is also given. The semantic classes have been separated quite well by acoustic content, with classical music concentrated on the right part of the map, the most quiet pieces located in the lower right corner. The musical opposite in terms of aggressiveness is found on the upper left corner: metal and punk music, followed by rock and pop beneath it and electronic music in the lower left. Both world and jazz music are located in-between classical music and the more energetic musical genres.

The SOM trained for the Latin music database (Fig. 5) was able to separate the genres even better, supposedly due to very distinct (and rather defined) characteristics in the different Latin American dances. Especially Axé, Bachata and Tango are grouped into almost pure clusters, but also the remaining genres are recognizable as cluster structures, although slightly interweaved.

We can produce multiple views for the SOM trained on the African music database, as different meta-data labels are available. The visualization in Fig. 6 shows the arrangement by country, where we see that the music from Congo DRC and Rwanda is separated on a coarse level (also with partial interleaving), Ghana forms a small cluster, and the less represented countries are aligned on the left edge, with Senegal placed above the Republic of Congo.

Fig. 7 shows the same alignment with the view of instrument families (where pieces with multiple instruments are indicated as separate classes). Although the map seems quite unstructured at first sight, there are some clusters of idiophone instruments, or pieces with chordophone+idiophone instruments. For a better clustering by instruments, however, a dedicated instrument detector should be used as the underlying feature extractor.

Generally speaking, a SOM can give insight into the inherent structure of music depending on the features extracted, and provides multiple views on a collection of music with different visualization metaphors. Especially

[6] White units with numbers represent songs with no class label available. Empty units were not populated by the SOM.

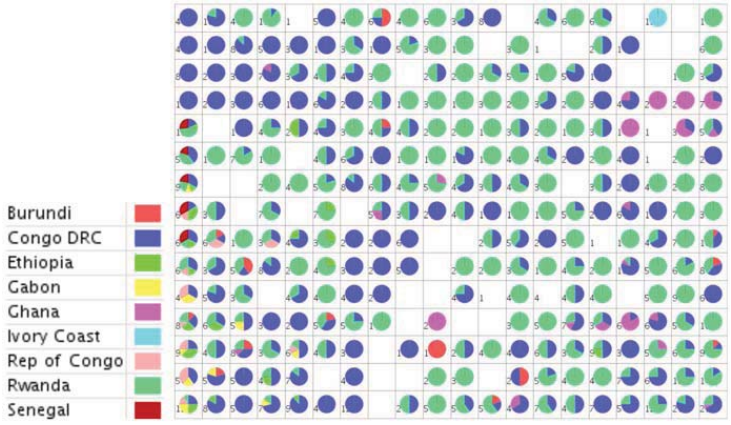


Fig. 6. Map of African music database, visualized by country.

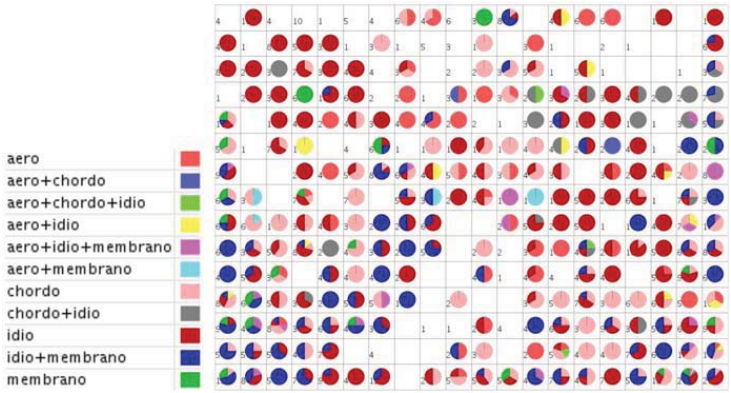


Fig. 7. Map of African music database, visualized by instrument family.

these multiple views and variable forms of visualization make the SOM (and the PlaySOM application) such a valuable tool for exploring ethnic music collections.

7. Conclusions

Improving means of access is essential to unlock the value that the holdings in ethnic, folk and other non-Western music collections represent. This includes tools to assist in analyzing, structuring and comparing the audio content of archives. These may help researchers in understanding complex relationships between the various pieces and assist in research work. They may also prove an invaluable asset when it comes to managing the increasing amounts of audio being digitized. Such tools, while not primarily geared towards research use, may also enable a broader public to get in contact with the massive volumes of valuable and rich cultural heritage recordings, such as of Irish or Greek folk music, Indian classical music or ethnic African music, familiarizing a larger public with this music. Yet, while a range of technical solutions are being developed in the field of Music IR the majority of these are designed, optimized and evaluated predominantly on Western music. Considering the peculiarities of non-Western and in particular ethnic music, both in terms of musical content and recording characteristics, the generality of the techniques developed needs to be considered. We conducted an in-depth analysis of the performance of a number of state-of-the-art and novel music analysis and audio feature extraction techniques on both Western and non-Western music. Their performance was evaluated on a range of classification tasks using machine learning techniques to structure music into predefined categories. Results were presented for three different music collections, specifically a benchmark collection with predominantly Western music, a database of studio recordings of Latin American music, as well as archival holdings of African music. Overall, the approaches proved to work surprisingly well in all these different settings. Major performance differences can rather be related to different musical characteristics (i.e. dominance in rhythm or timbre) rather than recording settings. It has been shown that state-of-the-art Music IR methods are capable to categorize an ethnic music collection also by meta-data such as the country or ethnic group, while the function of songs, an important attribute for ethnic music—by contrast to the genre used commonly for Western music—could not be recognized accordingly. Another finding is that ethnic music seems to be less susceptible to lead-in/fade-out effects and feature analysis delivers comparable results also from of the beginning of a

piece, reducing the effort for segmentation prior to audio feature extraction.

The second major contribution is the evaluation of a SOM-based interface to access recordings of non-Western and ethnic music. While this interface was predominantly developed as an access interface for playlist creation and the intuitive exploration of music collections, we demonstrated in this paper that it may also serve as a useful tool for more structural exploration of audio content. Relationships between various musical characteristics can be visualized and set in relation to each other, while at the same time serving as a convenient interface for the general public, who may lack the necessary in-depth knowledge to understand more traditional musicological organization schemes.

While the audio feature extraction and categorization approaches as well as structuring and access support have proven to work as well for non-Western and ethnic music recordings, the experiments suggest there is room for improvements. Similar to fine-tuning approaches specifically to Western music, dedicated modules considering musical structure in different types of non-Western recordings as well as pre-processing modules to account for different means of recording the audio seem necessary if evaluation across collections should be supported. Furthermore, adaptation of music analysis techniques to the specific characteristics of such music collections from the musicological point of view should be considered. In addition, efforts to analyze the textual content of non-Western and ethnic recordings in the vast range of languages found in such collections are essential to further close the semantic gap between the pure acoustic impression and the intention or function of specific pieces of music.

Acknowledgments

The authors would like to thank the Royal Museum of Central-Africa (RMCA), Belgium, for providing the data set of African music. We also would like to thank the Capes Brazilian Research Agency (process number 4871-06-5) and Mr. Breno Moiana for his invaluable help with the experiment infrastructure. We thank the reviewers of this article for helping to improve it through their thorough review.

References

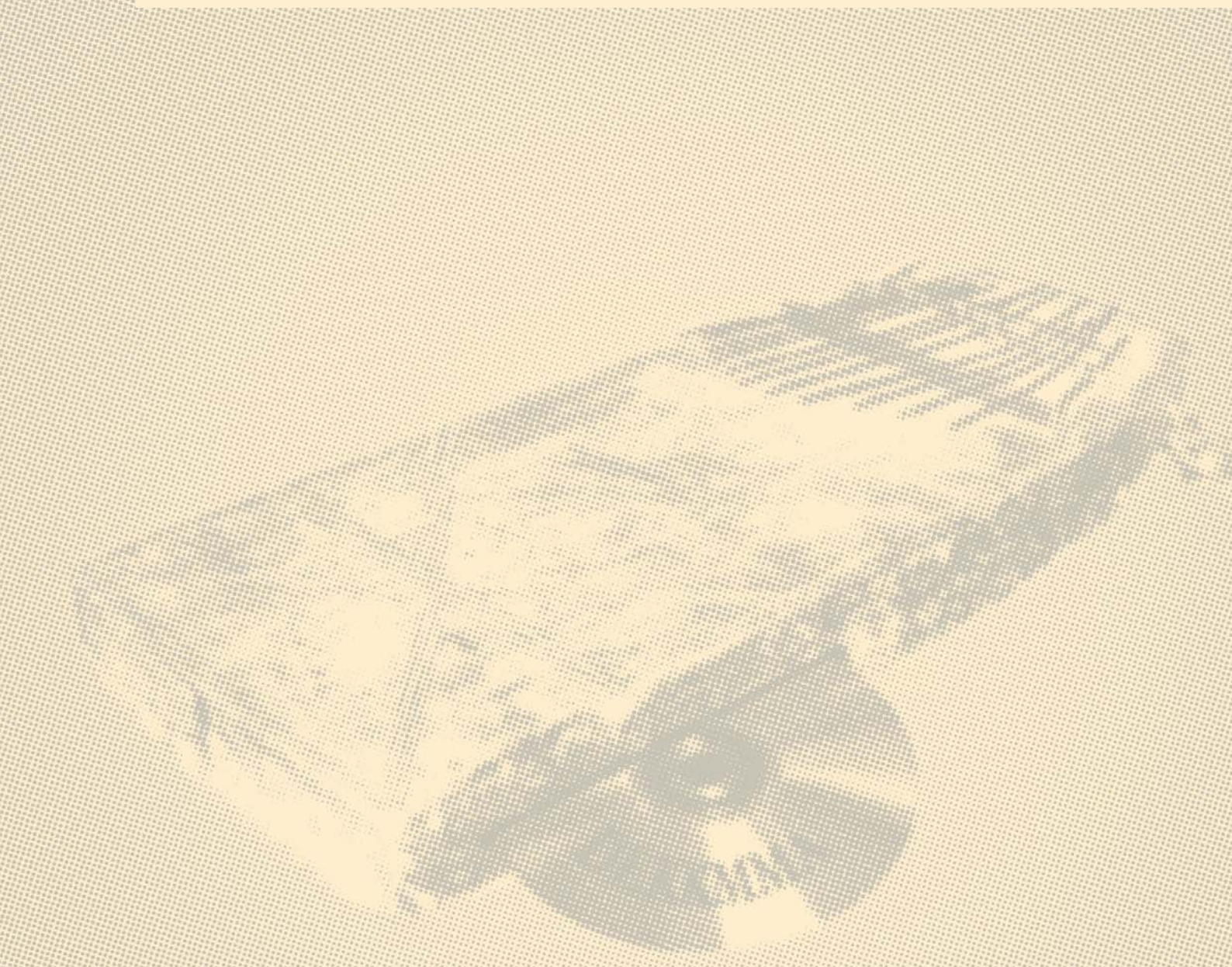
[1] N. Bernardini, X. Serra, M. Leman, G. Widmer, G. De Poli (Eds.), A Roadmap for Sound and Music Computing, The S2S2 Consortium, 2007.
[2] G. Tzanetakis, A. Kapur, A. Schloss, M. Wright, Computational ethnomusicology, Journal of Interdisciplinary Music Studies 1 (2) (2006) 1–24.
[3] M. Wright, A. Schloss, G. Tzanetakis, Analyzing Afro-Cuban rhythm using rotation-aware Clave template matching with dynamic programming, in: Proceedings of the International Conference on Music Information Retrieval, Philadelphia, PA, USA, 2008, pp. 647–652.
[4] F. Gouyon, Microtiming in "Samba de Roda"—preliminary experiments with polyphonic audio, in: Proceedings of the Brazilian Symposium on Computer Music, 2007, pp. 197–203.
[5] D. Moelants, O. Cornelis, M. Leman, J. Gansemans, R. De Caluwe, G. De Tré, T. Matthé, A. Hallez, The problems and opportunities of content-based analysis and description of ethnic music, International Journal of Intangible Heritage 2 (2007) 57–68.
[6] L. Auber, The Music of the Other, Ashgate Publishing, 2007.
[7] J.S. Downie, S.J. Cunningham, Toward a theory of music information retrieval queries: system design implications, in: Proceedings of the International Conference on Music Information Retrieval, Paris, France, 2002, pp. 299–300.
[8] J.-J. Aucouturier, F. Pachet, Representing musical genre: a state of the art, Journal of New Music Research 32 (1) (2003) 83–93.
[9] C. McKay, I. Fujinaga, Musical genre classification: Is it worth pursuing and how can it be, in: Proceedings of the International Conference on Music Information Retrieval, Victoria, Canada, 2006, pp. 101–106.
[10] G. Tzanetakis, P. Cook, Musical genre classification of audio signals, IEEE Transactions on Speech and Audio Processing 10 (5) (2002) 293–302.
[11] C. McKay, I. Fujinaga, Automatic genre classification using large high-level musical feature sets, in: Proceedings of the International Conference on Music Information Retrieval, Barcelona, Spain, 2004, pp. 525–530.
[12] R. Neumayer, A. Rauber, Multimodal analysis of text and audio features for music information retrieval, in: Multimodal Processing and Interaction: Audio, Video, Text, Springer, Berlin, Heidelberg, 2008.
[13] J.L. Herlocker, J.A. Konstan, L.G. Terveen, J.T. Riedl, Evaluating collaborative filtering recommender systems, ACM Transactions on Information Systems 22 (1) (2004) 5–53.
[14] X. Hu, J.S. Downie, K. West, A. Ehmann, Mining music reviews: promising preliminary results, in: Proceedings of the International Conference on Music Information Retrieval, London, UK, 2005, pp. 536–539.
[15] M. Schedl, P. Knees, T. Pohle, G. Widmer, Towards an automatically generated music information system via web content mining, in: European Conference on Information Retrieval, Glasgow, Scotland, 2008, pp. 585–590.
[16] P. Lamere, Social tagging and music information retrieval, Journal of New Music Research 37 (2) (2008) 101–114.
[17] T. Lidy, A. Rauber, A. Pertusa, J.M. Inesta, Improving genre classification by combination of audio and symbolic descriptors using a transcription system, in: Proceedings of the International Conference on Music Information Retrieval, Vienna, Austria, 2007, pp. 23–27.
[18] C. McKay, I. Fujinaga, Combining features extracted from audio, symbolic and cultural sources, in: Proceedings of the International Conference on Music Information Retrieval, Philadelphia, PA, USA, 2008, pp. 597–602.
[19] R. Mayer, R. Neumayer, A. Rauber, Combination of audio and lyrics features for genre classification in digital audio collections, in: Proceedings of the ACM International Conference on Multimedia, Vancouver, Canada, 2008, pp. 159–168.
[20] I.A. Bolshakov, A. Gelbukh, Computational linguistics: models, resources, applications, IPN-UNAM-FCE, 2004.
[21] C. Seeger, An instantaneous music notator, Journal of the International Folk Music Council 3 (1951) 103–106.
[22] A. Nesbit, L. Hollenberg, A. Senyard, Towards automatic transcription of Australian aboriginal music, in: Proceedings of the International Conference on Music Information Retrieval, Barcelona, Spain, 2004.
[23] A. Krishnaswamy, Melodic atoms for transcribing carnatic music, in: Proceedings of the International Conference on Music Information Retrieval, Barcelona, Spain, 2004.
[24] P. Chordia, A. Rae, Raag recognition using pitch-class and pitch-class dyad distributions, in: Proceedings of the International Conference on Music Information Retrieval, Vienna, Austria, 2007, pp. 431–436.
[25] D. Moelants, O. Cornelis, M. Leman, J. Gansemans, R. De Caluwe, G. De Tré, T. Matthé, A. Hallez, Problems and opportunities of applying data- and audio-mining techniques to ethnic music, in: Proceedings of the International Conference on Music Information Retrieval, Victoria, Canada, 2006.
[26] B. Duggan, B. O'Shea, M. Gainza, P. Cunningham, Machine annotation of sets of traditional Irish dance tunes, in: Proceedings of the International Conference on Music Information Retrieval, Philadelphia, PA, USA, 2008, pp. 401–406.
[27] A. Pikrakis, I. Antonopoulos, S. Theodoridis, Music meter and tempo tracking from raw polyphonic audio, in: Proceedings of the International Conference on Music Information Retrieval, Barcelona, Spain, 2004.

[28] I. Antonopoulos, A. Pikrakis, S. Theodoridis, O. Cornelis, D. Moelants, M. Leman, Music retrieval by rhythmic similarity applied on Greek and African traditional music, in: Proceedings of the International Conference on Music Information Retrieval, Vienna, Austria, 2007, pp. 297–300.
[29] N.M. Norowi, S. Doraisamy, R. Wirza, Factors affecting automatic genre classification: an investigation incorporating non-Western musical forms, in: Proceedings of the International Conference on Music Information Retrieval, London, UK, 2005, pp. 13–20.
[30] S. Doraisamy, S. Golzari, N.M. Norowi, M.N.B. Sulaiman, N.I. Udzir, A study on feature selection and classification techniques for automatic genre classification of traditional malay music, in: Proceedings of the International Conference on Music Information Retrieval, Philadelphia, PA, USA, 2008, pp. 331–336.
[31] J.S. Downie, Music information retrieval, Annual Review of Information Science and Technology, Information Today, vol. 37, Medford, NJ, USA, 2003, pp. 295–340.
[32] F. Gouyon, S. Dixon, E. Pampalk, G. Widmer, Evaluating rhythmic descriptors for musical genre classification, in: Proceedings of the 25th International AES Conference, London, UK, 2004.
[33] A. Rauber, E. Pampalk, D. Merkl, Using psycho-acoustic models and self-organizing maps to create a hierarchical structuring of music by musical styles, in: Proceedings of the International Conference on Music Information Retrieval, Paris, France, 2002, pp. 71–80.
[34] T. Lidy, A. Rauber, Evaluation of feature extractors and psycho-acoustic transformations for music genre classification, in: Proceedings of the 6th International Conference on Music Information Retrieval, London, UK, 2005, pp. 34–41.
[35] F. Gouyon, P. Herrera, P. Cano, Pulse-dependent analyses of percussive music, in: Proceedings of the 22nd International AES Conference on Virtual, Synthetic and Entertainment Audio, Espoo, Finland, 2002.
[36] A. Rauber, E. Pampalk, D. Merkl, The SOM-enhanced JukeBox: organization and visualization of music collections based on perceptual models, Journal of New Music Research 32 (2) (2003) 193–210.
[37] E. Zwicker, H. Fastl, Psychoacoustics—Facts and Models, Springer Series of Information Sciences, vol. 22, Springer, Berlin, 1999.
[38] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.
[39] J.C. Platt, Fast Training of Support Vector Machines using Sequential Minimal Optimization, MIT Press, Cambridge, MA, USA, 1999.
[40] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (3) (1998) 226–239.
[41] C.N. Silla Jr., A.L. Koerich, C.A.A. Kaestner, A machine learning approach to automatic music genre classification, Journal of the Brazilian Computer Society 14 (3) (2008) 7–18.
[42] C.N. Silla Jr., C.A.A. Kaestner, A.L. Koerich, Automatic music genre classification using ensemble of classifiers, in: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Montreal, Canada, 2007, pp. 1687–1692.
[43] P. Cano, E. Gómez, F. Gouyon, P. Herrera, M. Koppenberger, B. Ong, X. Serra, S. Streich, N. Wack, ISMIR 2004 audio description contest, Technical Report MTG-TR-2006-02, Music Technology Group, Pompeu Fabra University, 2006.
[44] ISMIR 2004 Audio Description Contest, Website, URL 〈http://ismir2004.ismir.net/ISMIR_Contest.html〉, 2004.
[45] C.N. Silla Jr., A.L. Koerich, C.A.A. Kaestner, The Latin music database, in: Proceedings of the International Conference on Music Information Retrieval, Philadelphia, PA, USA, 2008, pp. 451–456.
[46] O. Cornelis, R. De Caluwe, G. De Tré, A. Hallez, M. Leman, T. Matthé, D. Moelants, J. Gansemans, Digitisation of the ethnomusicological sound archive of the Royal Museum for Central Africa (Belgium), International Association of Sound and Audiovisual Archives Journal 26 (2005) 35–43.
[47] T. Matthé, G. De Tré, A. Hallez, R. De Caluwe, M. Leman, O. Cornelis, D. Moelants, J. Gansemans, A framework for flexible querying and mining of musical audio archives, in: Proceedings of the 16th International Conference on Database and Expert Systems Applications, 2005, pp. 1041–1045.
[48] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, second ed., Morgan Kaufmann, San Francisco, 2005.
[49] T. Kohonen, Self-organizing Maps, in: Springer Series in Information Sciences, vol. 30, third ed., Springer, Berlin, 2001.
[50] T. Lidy, A. Rauber, Classification and clustering of music for novel music access applications, cognitive technologies, in: Machine Learning Techniques for Multimedia, Springer, Berlin, Heidelberg, 2008, pp. 249–285.
[51] E. Pampalk, A. Rauber, D. Merkl, Using smoothed data histograms for cluster visualization in self-organizing maps, in: Proceedings of the International Conference on Neural Networks, Springer, Madrid, Spain, 2002, pp. 871–876.
[52] J. Frank, T. Lidy, P. Hlavac, A. Rauber, Map-based music interfaces for mobile devices, in: Proceedings of the ACM International Conference on Multimedia, Vancouver, Canada, 2008.

Moelants, D., Cornelis, O., Leman, M., Gansemans, J., Matthé, T., de Caluwe, R., De Tré, G., Matthé, T., & Hallez, A. (2007). Problems and opportunities of applying data- and audio-mining techniques to ethnic music. Journal of Intangible Heritage, 2 pp. 57–69.

The Problems and Opportunities of Content-based Analysis and Description of Ethnic Music

Dirk Moelants, Olmo Cornelis & Marc Leman
Jos Gansemans
Rita De Caluwe, Guy De Tré, Tom Matthé & Axel Hallez



The Problems and Opportunities of Content-based Analysis and Description of Ethnic Music

Dirk Moelants, Olmo Cornelis & Marc Leman
IPEM - Department of Musicology, University of Ghent, Belgium

Jos Gansemans
Department of Cultural Anthropology, Royal Museum of Central Africa, Belgium

Rita De Caluwe, Guy De Tré, Tom Matthé & Axel Hallez
TELIN, University of Ghent, Belgium

ABSTRACT

The Belgian Royal Museum for Central-Africa (RMCA) holds a large collection of ethnographic artifacts, including a sound archive with music recordings from the early 20th Century up to recently. The archive is one of the biggest and best-documented archives worldwide for the region of Central Africa. An on-going digitisation project is part of a strategy to conserve this archive and make it accessible to the public by (i) the digitisation of the data, and (ii) the application of music information retrieval techniques for the digitised data. While state-of-the-art research in music information retrieval aims to search and retrieve music on the basis of content description, most of the existing tools are designed for Western music collections, without any guarantee that these techniques can be applied to music from other cultures. African music, in particular, creates new challenges for content-based description and information retrieval. This paper describes some general problems regarding the content-based description of African and other non-Western music. It suggests an approach for describing pitch structures which will allow for the description of both Western music and non-Western music.

Introduction

During the last decade, digitisation projects for cultural heritage have received increasing financial support from both public sources and private institutions. This underlines the value and importance of preserving collections and making it easier to consult them by means of modern digital infrastructures and appropriate content management tools. Digital audio files can be

stored on a redundant storage system, which combines hard disks with a tape backup system for extra security. Nowadays, the technological infrastructure, based on digital broadband and mobile communication, makes it possible to access musical information quickly. Music can now be available from anywhere and at any time, by a finger click on the computer mouse, and there is no reason why this facility should be restricted to commercial music. The opportunities for storage, back-up, accessibility and documentation, make digital systems particularly well-suited to the preservation of cultural heritage. Unlike the playback of analogue audio, playbacks of digital audio do not harm the original carrier, and the number of copies cannot be exhausted. Digital meta-data, which typically consist of records with information about the origin and nature of the sound recording, and perhaps also the recording conditions, can be added to the musical audio files.

During the last decade, traditional meta-data descriptions which relate to the recording context have been complemented with another type of meta-description that is focused on the musical content. An objective content-description would typically focus on aspects such as timbre, pitch, melody, harmony, rhythm and tempo, while a subjective description would typically focus on factors related to movement (static, dynamic, slow, fast, etc.), emotional descriptors (gay, sad, etc.) or other semantic or corporeal descriptors (Lesaffre et al., 2004; Lesaffre, 2005; Leman, 2007).

So-called 'audio-mining' techniques aim to extract these content-based descriptions directly from the musical audio files. These techniques are based on low-level feature extraction and classification into higher-level descriptors that can be accessed by the human mind. For example, melody extraction from polyphonic audio is based on frequency analysis techniques that work on small time frames (e.g. 40 ms). The task is to select the fundamental frequency component that is representative for the melody, and to discard the information from percussion and accompanying instruments. At a higher analysis level, the time frames will be concatenated to form pitch objects that can be represented as an electronic score (e.g. Paiva, 2006). The melody provides a level of representation which users can access. For example, they can sing a melody and the sung melody can then be compared with a database of melodies that have been extracted from the audio archive.

The audio-mining approach is a first step in what is generally called 'data-mining' or the automatic search for patterns in large volumes of data, using techniques of statistical correlation and categorisation. Indeed, users would typically tend to extract further information from such a database, such as information about similarities in a large collection of musical pieces. Similarities can then be represented graphically, on maps and visualisation schemas (van Gulik & Vignoli, 2005; Pampalk 2005). Music information retrieval thus aims to combine audio-mining techniques and data-mining techniques for the search and retrieval of music in a digital music library. This would allow new query techniques to be developed - like searching for similar pieces of music (including the possibility of 'query-by-example', where the user uploads his audio fragment) or the use of semantic descriptions (e.g. adjectives relating to emotions or gestures) that are automatically connected to the low-level features of the musical content (Leman et al., 2005). However, although the techniques for content-based description and retrieval look promising, it remains very difficult to get technological success rates that would allow practical, real-world applications. There is a huge semantic gap between music as digital encoded audio, and music as

something meaningful. The step from encoded audio to meaning involves many aspects of human perception, understanding and a thorough knowledge of the social context in which the music information retrieval activities are taking place.

At this moment, the audio-mining techniques that have been developed are designed principally to extract information from Western music. Hence, the concepts that underlie audio-mining techniques are often based on Western music theory. Examples are the use of the chromatic (equidistant 12-tone) scale or the assumption of a regular division of the measure in ternary or binary units. There seems to be a general lack of knowledge about cultural heritage from beyond the Western world, and music that has a fundamentally different structure is not usually considered at all.

The construction of a database with 'traditional' meta-data for non-Western, and especially ethnic, music requires an approach that is entirely different from that used for Western popular and classical music. In this paper we describe some of the problems encountered during the process of digitising the music archives of the Belgian Royal Museum of Central-Africa (RMCA), and we propose an approach that would allow more suitable descriptions for music with structural and sociological characteristics that are very different from the Western standard. First, we will give a short description of the collections of the RMCA and the digitisation process. Then we will introduce some of the problems - and opportunities - we have encountered in dealing with databases of non - Western music, and in the last section we will present a system that analyses and compares the pitch content of musical pieces as an illustration of how content-based descriptions of music can deal with specific aspects of non-Western music.

Digitisation of the audio collection of the Belgian Royal Museum of Central-Africa (RMCA)

With its 50,000 sound recordings (with a total of 3,000 hours of music), dating from the early 20th century up to the present day, the music archive of the RMCA (located in Tervuren near Brussels) is one of the biggest in the world for the region of Central Africa. The music archive is part of a larger collection of artifacts from Central Africa that includes musical instruments, masks, tools, animals, plants and many other objects. Conservation and access to this collection is of particular importance given the current political instability, and the general destruction of cultural heritage that is taking place in this region (Cornelis et al., 2005). To conserve this important cultural heritage, and make it accessible to the public, the Belgian government provided a grant for a project called DEKKMMA. The project aimed to (i) digitise the different types of sound recordings (wax cylinders, sonofil, vinyl recordings and magnetic tapes) each with their own specific problems, (ii) construct a database suitable for describing different types of music (Matthé et al., 2005), (iii) explore tools that will allow further analysis

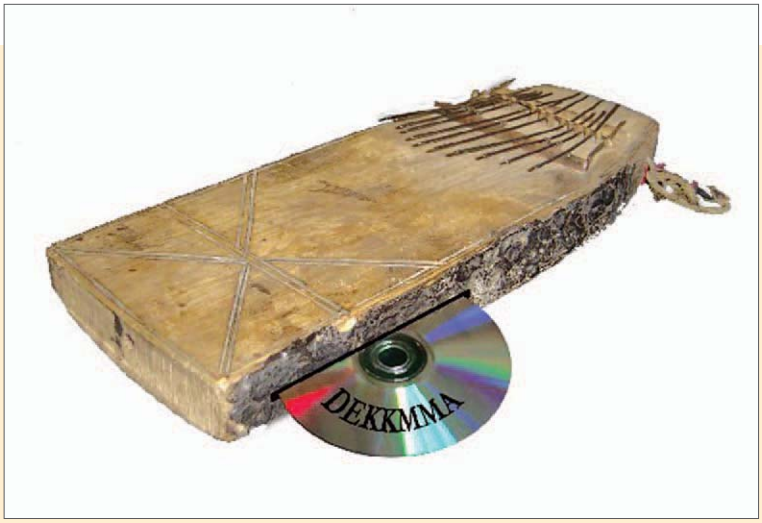


Figure 1
Logo of the DEKKMMA-project, symbolising the digitisation of the African music archive

and classification using audio-mining and data-mining techniques, (iv) develop content pertaining to the description of musical culture and musical instruments from different geographic regions, (v) process audio, photographic and video material from the archives and from the museum collections as background information for the audio files. (The results of this project can be accessed on the website http://music.africamuseum.be/. The reader will find metadata as well as sound excerpts and accompanying documentation such as descriptions of musical instruments.)

The project included the construction of the website, which allows different user groups to search and retrieve data related to the music archive. Three user groups were identified. The first and largest group are people who are just interested in African music, but do not have much knowledge of it. These users typically want to retrieve music using a rather vague and general labelling, such as 'drumming', 'trance music' or 'some song from Rwanda'. The second group consists of users from Central Africa. They often have a good knowledge of certain repertoires and functions of the music, and therefore they tend to ask very specific questions - such as for music played by a specific performer, music from one particular village, lyrics, genres, instruments - and they may well use local terminology. Finally, the third group of users consists of researchers who use the database for further study. This group would typically tend to ask questions related to the geographical spread of certain types of instrument, or the relative importance of certain rhythmic or pitch musical structures in different regions. In short, music information retrieval has to take into account user groups with a whole range of different interests. Research in music information retrieval aims to develop new tools and find proper ways of dealing with the interests of these different user groups. In practice, this often requires a thorough empirical analysis of the needs, as well as the subject backgrounds, of the users (Lesaffre et al., in press).

To overcome the lack of knowledge of the amateur user, extra search strategies have been implemented, such as searching by clicking on a map of Africa, or listing musical instruments in a tree according to Sachs' and von Hornbostel's musical instrument classification. For example, the tree structure allows the user to search for wind instruments, then for flutes, then for straight and transverse flutes and finally a list of vernacular instrument names will be given, including the names of the countries where a particular instrument has been found. Interesting applications to be developed for the first group of users are query-by-example, or search by effective parameters. Query-by-example would allow users to provide an audio example and formulate a request to find music with a similar rhythm. Query by effective parameters would allow users to specify music using descriptors that pertain to particular emotions, musical effects, moods, or movement characteristics. Users can also add valuable information to the database. For this reason, a forum will be set up where users can discuss topics, improve the descriptions of content and, hopefully, fill some gaps. The forum will allow users to come into contact with researchers working on developing the database, and that may produce ideas for the invention of new search tools and possibly allow the database to be extended to include copies of recordings kept at other institutions or by private individuals. In DEKKMMA, some of these facilities have been implemented, while other parts are still at the development stage.

Integrating non-Western music into databases: problems and opportunities

The major task of the DEKKMMA project is to transform the (analogue) music archive into a digital music library, using modern tools of music information retrieval. However, music information retrieval research usually takes Western music and its musical characteristics and semantic descriptions as a standard, and develops tools following a series of assumptions based on Western cultural concepts. These apply to structural aspects (e.g. tonal key, assumption of octave equivalence, instrumentation), social organisation of the music (e.g. composers, performers, audience) and technical aspects (e.g. record company, release date). There is no guarantee that these concepts can be readily applied to non-Western music. Indeed, the production and appreciation of music in oral cultures may be completely different from the way Westerners see music, with traditions of learning by listening and practicing, passing skills from father to son or from master to pupil. Trying to incorporate descriptions of non-Western music in databases that are structured to cope with the demands of classifying Western music, often causes problems. These problems can be found both in descriptive meta-data and in descriptions of musical content. Imposing

Western concepts on to non-Western music can lead to incorrect information going into the databases, or important information being excluded through the lack of suitable database fields.

Music Information Retrieval (MIR) applications that focus on popular and classical Western music will make the search and retrieval of such music easier, and therefore more people will have access to it. In contrast, music that occupies a more marginal position risks being excluded by this technology, and will thus become even less accessible. In this way, the combination of digitisation and commercial large-scale distribution tends to push 'vulnerable' music even further into oblivion. Music information retrieval research should therefore take into account an ethical code that aims to develop tools for all types of music, not just for Western music. This is a huge challenge for music research and it will require a change of approach in most centres that currently deal with music information retrieval. To bring people into contact with music they would normally never have heard of requires a reconsideration of the concepts that underlie musical practices in non-Western cultures.

Simply integrating more ethnic music, and non-Western music in general, into the existing databases and indexes will solve the above-mentioned problem entirely. Indeed, as already mentioned, musical structures, as well as the relative importance of structural elements within the musical experience, can be fundamentally different in different cultures. A straightforward example is the Western focus on pitch and fixed tuning, whereas in African music fixed tuning does not exist. Instead, a large number of different pitch scales can be observed, and often relative pitch (higher-lower) is more important than absolute pitch. A proposal for a method that deals with the description of pitch will be outlined in the next section.

Another difficulty with integrating ethnic music into digital music libraries concerns the organisation of the meta-data. In describing field recordings, it often happens that some information that is 'compulsory' in the description of Western music is lacking, while other information that seems irrelevant in the description of Western music turns out to be very important. For example, names of composers are usually not known. Performers could be named but music is often seen as performed by 'the community' and therefore, the names of the participants are not considered to be very important. On the other hand, the location and date of the recording are important because location can be a crucial search field for retrieving music from oral cultures. Since names of performers and composers are usually not known, the music is primarily identified with the country, region, ethnicity or town where it is produced. In many existing databases, location is not even an existing meta-data field, due to its low relevance in Western music.

A further problem in the meta-data descriptions of field recordings is related to the lack of standardisation. This is due to the fact that these meta-data descriptions often have a historical origin, and have been collected by many different field researchers, often amateurs, who used many different recording techniques. As a result, not all recordings are equally well-documented. For some interesting old field recordings even the most basic information is lacking, and one needs extensive historical research in order to have even a rough idea of the time and place where the music might have been recorded. But even recordings made by professional ethno-musicologists are sometimes not completely documented. In some cases, the documentation has been partially lost, or the connection between the recordings and the documentation is no longer clear. Given the fact that knowledge about traditional music within oral cultures is vanishing under pressure of urbanisation and Westernisation, the correct identification of the music and its meta-data descriptions, as well as the definition of its authenticity (in the digital context) becomes increasingly important.

Finally, there is a problem of terminology. There can be different local names for the same concept, and different researchers can use different terms for them. At this moment, the American Folklore Society and the American Folklife Center at the Library of Congress are constructing an 'Ethnographic Thesaurus', a comprehensive, controlled list of subject terms to be used in describing ethnographic and ethnological research collections (cf. http://www.etproject.org). But even a standardised list cannot solve all the problems. Consider the example of the 'thumb piano' (lamellophone). This instrument type has very diverse names (see Table 1). A user looking for one of these names should also be directed to pieces in which one of the other terms is used. This requires an elaborate thesaurus and a specific approach in the construction of the database (Matthé *et al.*, 2006). To make it even more complicated, one name does not necessarily point to a specific sub-type: size, material, number of pitches and

Table 1
Variable denominations for the lamellophone.

Tshisaasj	Ngombi	Kadimba	Agidogo
Tshisaji	Kombi	Kalimbe	Alogu
Tshisanji	Kembe	Malimba	Ashwa
Chisanzi	Ekembe	Marimba	Dongo
Sanzi	Ikembe	Irimba	Isen
Sanza	Likembe	Ilimba	Kankobele
Sansa	Dikembe	Limba	Mambalat
Esanzo	Kidebe		Prempensua
Issanji	Ibekee		Ubo
Kisanzi	Gbelee		· · ·
Kassandji			
Kassayo			

tuning can vary widely. Therefore it is desirable that the user should be able to refine his search by looking for more specific instrument characteristics, or for instruments with similar tuning.

Audio-mining and the pitch structures in African music

Research on content-based music information retrieval aims i) to define the search and retrieval of music in terms of musical content descriptors and ii) to develop automated content-description and retrieval methods. Rather than having to specify the name of the composer or the title of the song, the content-based approach would allow one to specify musical content using descriptors related to its nature such as `happy', 'sad', 'dynamic', 'harmonious', or using corporeal descriptors which define particular movements that are captured by sensors (e.g. indicating tempo or expression), using graphical navigation in databases, and using search and retrieval by providing audio examples (for relevant publications, see www.ismir.net). As content-description is often based on subjective descriptions, knowledge of the user's background is an important issue (Lesaffre et al., in press).

In this section, we focus on one single structural characteristic of musical content, namely pitch. Although pitch is closely related to the physical characteristics of music, and less to subjective factors that would depend on education, gender, and familiarity, there are many problems with this as a content descriptor. The major problem is that researchers tend to develop pitch extraction algorithms that are based on Western concepts and assumptions from music theory which cannot easily be applied to non-Western music. A straightforward example of such an assumption is the concept of the octave-reduced pitch representation, with a categorisation based on the chromatic 12-tone scale.

Within the collection of Central African music of the RMCA, there are a wide variety of tunings and scales. Often, a melody is closely connected to the tones of the Bantu tone languages. This is the case in vocal music where the melody-line has to follow the speech-tones, and in instrumental genres that are based on verbal elements. A similar phenomenon is seen in the music of other tonal languages, such as Chinese. In Bantu languages there is a continuum between speech and music, which makes it possible to transfer messages from one village to another by using drum signals of different pitch. The use of speech tones that do not have a precisely determined pitch makes it more important to distinguish between high and low pitches rather than having specific harmonic relationships between pitches. Typically, in this region, people prefer instrumental sounds with a very broad, percussive spectrum, which in turn reduces the importance of 'correct' pitch intervals. However, other instruments with fixed tuning, like flutes, zither, (wooden) trumpets or thumb pianos, allow the study of possible fixed pitch relations and pitch scales.

To include information about pitch scales and pitch tunings in the DEKKMMA database, a pitch description approach was used that represents pitch scales without reference to Western notes or scales. In avoiding a priori pitch categories, this approach is based on a continuous representation of pitch. Thus, rather than a discrete pitch representation, we propose a continuous pitch representation and continuously-based associated retrieval mechanisms.

In the first stage, the music is analysed by a melody extractor (De Mulder et al., 2004). This melody extractor

Figure 2
Pitch analysis of Ingendo y'inka. The x-axis shows the pitch in cents (0 = a 55 Hz), the y-axis shows the number of occurrences for every 1 cent interval, smoothed by taking the average of 7 bins around the middle. The vertical lines show the position of the a's, dividing the whole into three octaves.

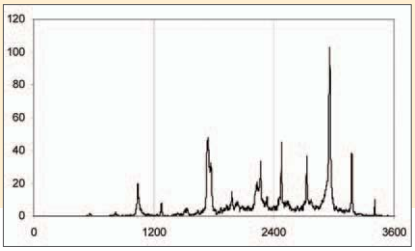


Figure 3
Octave-reduced pitch analysis of Ingendo y'inka. The x-axis shows the pitch with 0 = a, the y-axis shows the number of occurrences for every 1 cent interval, smoothed by taking the average of 7 bins around the middle. The vertical lines show the position of the Western standard pitches, every 100 cents.

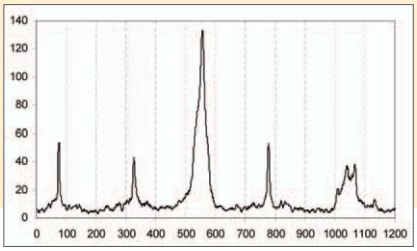


Table 2
Analysis of the most prominent peaks in Ingendo y'inka: the left column shows the pitch in cents above 55 Hz, the middle column shows the distance in cents between successive pitches and the right column shows the size of the octave relationships.

Pitch	interval	octave
1042	236	1225
1278	264	1199
1542	202	1185
1744	237	1214
1981	286	1198
2267	210	1140
2477	250	
2727	231	
2958	221	
3179	228	
3407		

was originally designed for the transcription of vocal queries. It was optimised for monophonic music and the normal voice range. Yet, testing the model on different types of music reveals that it can provide a straightforward image of the pitch distribution, even in music with a more complex texture. For complex polyphonic music with a dense texture, like Western symphonic music, the system is less successful, but other pitch detection systems could be used, following the same methodology. The melody extractor currently used gives a frequency for every time frame of 10 ms. In order to give an image of the scale, we transform these values to cent values (taking the low A, 55Hz as 0 cents). The cent scale divides every half tone into 100 subdivisions, which allows a very precise representation of the actual pitch. An additional advantage of the cent-scale is that the distances are the same in every octave and within every half-tone (this is not the case in a Hertz frequency representation, which has a somewhat logarithmic relationship to pitch). The representation of the pitch content of a piece is then given by plotting the number of occurrences of each cent value between 1 and 6000 cents. In this paper the method will be illustrated first by the analysis of one particular song, then seven other examples, some from the same region and some from completely different musical cultures, will be used to show how this methodology can be applied to a broad spectrum of types of music.

The sample song is called *Ingendo y'inka*. It is performed by a man singing with *ikembe* (thumb piano) accompaniment. The piece was recorded on 14 January 1973 by Jos Gansemans in the village Karengera, Cyangugu province, Rwanda. The singing is mainly in *parlando* style, following the pitches of the instrument. The text describes the elegance of the local cows, because these animals are very importance in local society and are a sign of wealth. Extraction of the pitches in this song reveals that 11 notes occur, spread over several octaves (see figure 2, table 2). The *tessitura* used is quite low; within the octaves A to a' (110-440 Hz). Certain tones occur in the different octaves. A reduction to one octave of 1200 cents gives a better picture of the scale used, but at the same time it illustrates the danger of this octave reduction. Figure 3 shows the octave-reduced representation. The subdivisions on the X-axis (the discontinuous vertical lines) represent Western equal-tempered tuning: Every half tone has a distance of 100 cents, a whole tone measures 200 cents, a small third 300, a fifth 700, the octave 1200. These markers give a clear picture of the pitches of the song, and in this case put forward only 5 different notes that are frequently used. Looking at this pentatonic scale, 3 intervals around 226 cents and two around 266 cents (table 2) can be found.

The African tonality in this case, as measured, can be described as more or less equidistant pentatonic, although the small differences in interval size might be characteristic for this scale. The peak between 1000 and 1100 cents (g and g#) is much broader, and we can, in fact, see three small peaks in it (Figure 3). Although in most music the octaves and fifths relations are usually quite precise, due to the strong sense of consonance, we see that is not the case in this example. The highest octave is much (60 cents) too small, while the lowest is 25 cents too wide. These deviations are not uncommon in African music and they create a sought-after tension

Figure 4 : 73.9.3-6, flute
Octave-reduced pitch analysis of an umwirongi(flute) piece from Cyangugu province, Rwanda.

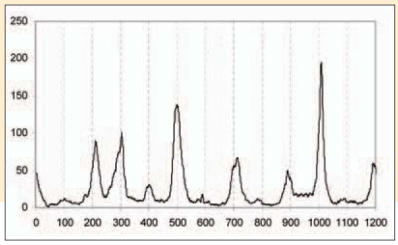


Figure 5 : 73.9.15-10, flute
Octave-reduced pitch analysis of an umwirongi(flute) piece from Ruhengeri province, Rwanda.

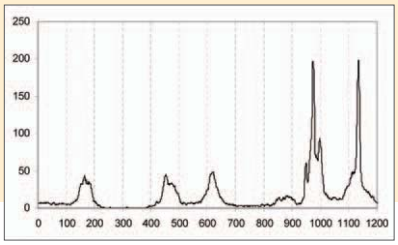
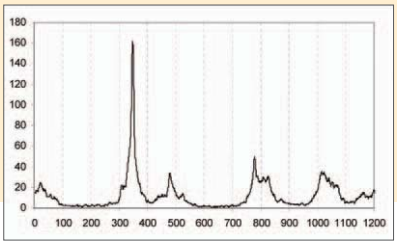


Figure 6 : 54.1.13-1, musical bow and singing
Octave-reduced pitch analysis of a Twa song with umuduri (musical bow) accompaniment from Kigali province, Rwanda.



instead of smooth sounding exact octaves. This shows that the octave-reduction has both an advantage and a disadvantage. It makes it much easier to illustrate the scale and to compare it with other examples, but it removes the subtle differences between octaves that are important in some styles of music.

For six other pieces, an octave-reduced graph is created in order to show the use of this method in characterising and comparing different tuning systems and scales. Thus every peak indicates the number of annotations made for every tone. Three examples (Figures 4-6) are taken from the collection of Rwandan music in the RMCA. The first two were also recorded in 1973 by Jos Gansemans, and they are both played on the *umwirongi* flute. The first (Figure 4) was recorded in the village Cyimbogo in the same Cyangugu province as *Ingendo y'inka*, the second (Figure 5) in the village Ndusu in Ruhengeri province. The third example (Figure 6) represents the pitch content of a song accompanied by the musical bow (*umuduri*) from the Twa people in the village of Nyanza in Kigali province; it was recorded in 1954 by the missionary Scohy-Stroobants.

All the examples point to pentatonic tunings. However, there is no standard. Figure 4 is very similar to figure 3, using a more or less equidistant pentatonic scale. If the pitch of figure 4 were lowered by 330 cents, the main peaks would fall together, and the two would be almost identical. Although the second piece was played on the same type of instrument, the tuning is clearly different, tending more towards irregular anhemitonic pentatonic scales. Another difference is that instead of having one main peak, this example has two large peaks of almost equal magnitude. The third example stands somewhere in between, and could be related to the others, even though the origin, instrumentation and recording date are very different.

Three additional examples illustrate how different musical styles and scales are represented. Figure 7 shows the pitch distribution of Mozart's piano sonata in G Major (KV 283). Figure 8 shows a piece of Korean classical orchestral court music (*Pohoja* performed by the Seoul National Orchestra of Classical Music), and Figure 9 shows a piece of Persian santur (zither) music (*Avaz* from *Bayate Esfahan* played by Mohamad Heydari).

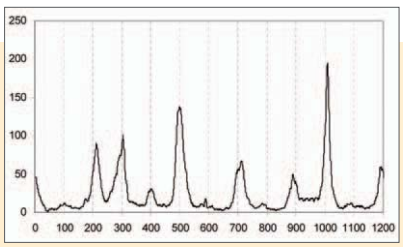


Figure 7 : Pianosonata Mozart
Octave-reduced pitch analysis of W.A. Mozart's pianosonata in G-major (KV 283)

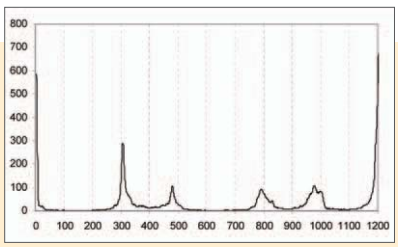


Figure 8 : Korea-Aag
Octave-reduced pitch analysis of *"Pohoja"*, banquet music for the Korean court, performed by the "Seoul National Orchestra of Classical Music".

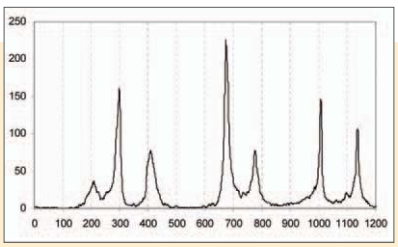


Figure 9 : Iran-Santur
Octave-reduced pitch analysis of *Avaz from Bayate Esfahan* perforemed by Mohamad Heydari on the Persian citer santur.

These examples illustrate the difference between distinct Western and non-Western tone systems very well. The Western system uses a chromatic/diatonic scale in which interval sizes are multipliers of 100 cents, and there is a strict adherence to a 1200 cents octave. The Mozart piece clearly shows its G-major tonality and 100/200 cents based structure (in all examples 0 cents is A, 100 cents is A#, 200 cents is B") so peaks appear at 1000/300/500 (representing notes G/C/D, which are the tonal grades I-IV-V in G-Major), and the peak at 200 cents (note B, III grade) reveals the major tonality.

In contrast, the Korean and the Persian pieces rely on well-defined tuning systems, but the graphs show characteristics that would not be represented idiomatically using Western tonal notation. The pitches used in the Korean court orchestra are closely related to the Western tuning system, but the music uses a pentatonic scale, just as in the African pieces described above, and the pitches do not correspond exactly to the divisions of the Western scale. The Persian example is clearly heptatonic, like most Western scales, but the distances between the pitches do not adhere to the standard patterns as the Persian modes use both 'normal' semi-tones and 'enlarged' intervals, like 3/4-tones. Comparisons of pitch scales can be based on continuous pitch representations using cross-correlation techniques. Although this is computationally intensive, it offers an appropriate method that avoids the danger of imposing Western pitch categories on to non-Western music. Further study may reveal that the cent-scale can be represented at a lower sampling rate to reduce the computational cost in search and retrieval. However, the concept of a continuous representation over different octaves would remain the core feature of this pitch representation schema.

Conclusion

Constructing applications for content-based descriptions of music that can deal with all the world's musical traditions is difficult, but this seems very necessary to protect the world's cultural heritage. It is important to bring together knowledge about the music from different cultures, and to make it accessible to a broad audience. Dealing with ethnic music reveals interpretational problems related to musical practice, the semantic description of musical features, as well as the automated extracted musical content parameters. The fundamentally different use of pitch in African music compared to Western music illustrates the difficulty of applying existing Western discreet pitch categories to non-Western music. Other examples pertain to aspects of rhythm, timbre and articulation. In addition, one should be careful about adopting extra-musical descriptive characteristics, which typically relate to musical practice in its social/cultural context. The use of our Western semantic and perceptual framework is often inappropriate for accurate digitalisation and the subsequent development of a digital library for cultural heritage.

The DEKKMMA-project is an example of a digitalisation project that aims to develop methods for content-based description for all types of music. This was illustrated by the method for representing pitch scales, using a continuous representational schema that can be used in search and retrieval applications.

REFERENCES

- Cornelis, O., De Caluwe, R., De Tré, G., Hallez, A., Leman, M., Matthé, T., Moelants, D. & Gansemans, J. 2005. Digitisation of the Ethnomusicological Sound Archive of the Royal Museum for Central Africa (Belgium). pp. 35-43 in *IASA journal*, 26.
- De Mulder, T., Martens, J.P., Lesaffre, M., Leman, M., De Baets, B. & De Meyer, H. 2004. Recent improvements of an auditory model based front-end for the transcription of vocal queries. pp. 257-260 in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. (Montreal)
- Leman, M. 2007. *Embodied Music Cognition*. (Cambridge: MIT Press)
- Leman, M., Vermeulen, V., De Voogdt, L., Moelants, D. & Lesaffre, M. 2005. Prediction of Musical Affect Using a Combination of Acoustic Structural Cues. pp. 39-68 in *Journal of New Music Research* 34(1)
- Lesaffre, M. 2005. *Music Information Retrieval: Conceptual Framework, Annotation and User Behavior*. (Unpublished Ph D. thesis: Ghent University)
- Lesaffre, M. Leman, M. De Baets, B. Martens, J.-P. 2004. Methodological considerations concerning manual annotation of musical audio in function of algorithm development. pp. 64-71 in *Proceedings of the fifth International Conference on Music Information Retrieval.* (Barcelona)Lesaffre, M., De
- Voogdt, L., & Leman, M. (2007) "How potential users of music search and retrieval systems describe the semantic quality of music" (in press)
- Matthé, T., De Tré, G., Hallez, A., De Caluwe, R., Leman, M., Cornelis, O., Moelants, D. &Gansemans, J. 2005. A Framework for Flexible Querying and Mining of Musical Audio Archives. pp.1041-1045 in *Proceedings of the 16th International Conference on Database and Expert Systems Applications*. (Copenhagen)
- Matthé, T., De Caluwe, R., De Tré, G., Hallez, A. Verstraete, J., Leman, M., Cornelis, O. Moelants, D. &. Gansemans, J. 2006. Similarity Between Multi-valued Thesaurus Attributes: Theory and Application in Multimedia Systems. pp. 331-342 *Lecture Notes in Computer Science 4027.*
- Paiva, R. P. 2006. *Melody Detection in Polyphonic Audio*. (Unpublished Ph.D thesis: University of Coimbra)
- Pampalk, E. 2004. A Matlab Toolbox to compute Music Similarity from Audio, pp. 254-257 in *Proceedings of the Fifth International Conference on Music Information Retrieval*. (Barcelona)
- Van Gulik, R., & Vignoli, F. 2005. Visual playlist generation on the artist map. pp.520-523 in *Proceedings of the Sixth International Conference on Music Information Retrieval*. (London)

Oramas, S., Cornelis, O. (2012). Past, present and future in ethnomusicology: the computational challenge. Proceedings of the 13th ISMIR Conference, October 8th-12th, 2012, Porto, Portugal.

PAST, PRESENT AND FUTURE IN ETHNOMUSICOLOGY: THE COMPUTATIONAL CHALLENGE

Sergio Oramas
Polytechnic University of Madrid, Spain
soramas@gmail.com

Olmo Cornelis
University College Ghent, Belgium
olmo.cornelis@hogent.be

ABSTRACT

Ethnomusicology has changed its paradigm over the years, but the core of this field is mainly related to non-Western and folk music studies. It combines an anthropological and musicological point of view, not only studying the sound itself, but also its context. The MIR community is evincing interest in non-Western traditions, but this interest is still very recent. During the late break session at ISMIR 2012 (Porto), several researchers joined the session on ethnomusicology and some ideas were proposed, sketched in this paper. First, it was suggested the necessity of creating a web site or wiki page for the publication of content related to the field. Second, it was stressed the importance of the creation of an international research network. Third, the connection between MIR researchers and ethnomusicologists was emphasized once more.

1. INTRODUCTION

1.1 Ethnomusicology

Ethnomusicology is an interdisciplinary field of study whose term was coined in 1950 by Jaap Kunst [1]. It was not suddenly developed, but has its roots in a previous scientific field with interest in non-Western and folk music called comparative musicology [2].

After some folklore studies done by composers like Béla Bartók or Zoltán Kodály, comparative musicology developed at the end of the 19th century in Vienna and Berlin by Carl Stumpf, Curt Sachs, Erich Von Hornbostel, and Otto Abraham. Its objective was the study and understanding of music by comparing different cultures with the aim of finding musical universalities and the origin of music. It was also the beginning of the creation of sound archives.

Around 1950, the evolution of these studies led to the creation of a new discipline called ethnomusicology. With this new concept researchers such as Alan P. Merriam or Bruno Nettl pursued to strengthen the link with anthropology [3] and also emphasize the importance of the focus point of the researcher rather than the object of study itself. Fieldwork became an essential part of any research and contextual information was considered crucial [4].

In the following decades, there was a confrontation between two different viewpoints, a more anthropological, sociological one and a more analytical, musicological one. Nowadays, ethnomusicologists are supposed to cover both perspectives within their research, integrating the study of sound and its cultural context.

Moreover, the interests of scholars have expanded significantly and today not only do they encompass sound events and their relationship to context, but also could include the study of their historical dimension, gender aspects associated with musical activity, iconographic representation, motor aspects related to production of sound and dance, the relationship between execution of musical instruments and the human body, and many other issues.

1.2 Computational Ethnomusicology

Since late 20th century more and more audio archives have been digitized. This process led to some (rather individual) initiatives to create automated musical content descriptions. Tzanetakis launched the name computational ethnomusicology [5], and since then interest in non-Western music research in the MIR community has grown [6]. Several MIR techniques have already been applied to the analysis of some ethnomusicological corpora, such as pattern matching, melodic similarity, music emotion recognition, and so on. Some research groups has been created to study these topics, such as the COFLA (COmputational analysis of FLAmenco music) group [7]. Moreover, a first large European funded project called CompMusic has been released, which focuses on Turkish, Chinese and Indian music [8].

2. NEW CHALLENGES

A new discipline always comes with new research questions, asks for new methodologies, and contains new challenges and problems.

2.1 Reflection: what to research, for who, and how!

Are current studies focused just on sound or also on the relationship to context? Are they fitting the needs of ethnomusicologists? Do the current studies provide useful and accessible tools? Is it easy to find information on the Web about these tools? Is the MIR community aware of the

last tendencies in ethnomusicology? All these questions should be posed by the MIR community in order to really understand issues and tendencies in computational ethnomusicology.

2.2 Pragmatical: being organized

Since a few years, an interest group on computational ethnomusicology has been organized through a mailing list. A modest attempt in receiving COST funding for creating an international network was rejected, despite positive and hopeful comments. Also, a new conference saw the light: FMA (Folk Music Analysis, held in Greece, Spain, and upcoming Holland (2013) and Belgium (2014)), where the community shows their new challenges and developments in the field.

2.3 Present issues

Current MIR research mainly focuses on Western music. The results obtained are not straightforwardly transferable to music of other cultures whose musical concepts that do not always correspond to the Western theoretical music concepts [9]. Therefore, it is important to discuss the reasons why existing techniques fail and what strategies are required. There is an urgent need for collaboration between musicologists and MIR researchers. Another recurring problem is the lack of ground truth. Such set might be attractive for MIR researchers in general (e.g. task in MIREX).

3. CONCLUSION

The interest and role of MIR in computational ethnomusicology is increasing, and a new interdisciplinary field is taking shape. But there is still a long way to go before we can speak of a well-established discipline. Our community has two main organizational challenges. The first is the creation of a central platform, such as a web site or a wiki page, where news, events, publications, data-sets, and tools could be posted and shared by researchers. A place where ethnomusicologist as wee, could find new tools to use in their work; where researchers in computational ethnomusicology could find datasets to test their algorithms or quickly see the state of art of a specific problem. The second challenge is the creation of an international network that will help in the development of the field, providing funding to support interchange of researchers and promotion of activities such as workshops and seminars, which will encourage the interest in and broaden the knowledge of computational ethnomusicology.

4. REFERENCES

[1] Kunst, J.: *Musicologica.* Royal Tropical Institute, Amsterdam, 1950.

[2] Adler, G.: *Umfang, Methode und Ziel der Musikwissenshaft.* Vierteljahrsschrift fr Musikwissenschaft 1, pp. 5–20, 1885.

[3] Merriam, A.P.: *The Anthropology of Music*. Evanston, Northwestern University Press, 1964.

[4] Nettl, B.: *Theory and Method in Ethnomusicology*. The Free Press of Glencoe, London and New York, 1964.

[5] Tzanetakis, G., et al.: "Computational ethnomusicology" *Journal of Interdisciplinary Music Studies 1 (2)*, pp. 124, 2007.

[6] Cornelis O., Moelants D., Leman M.: "Global access to ethnic Music: the next big challenge?" *Proceedings ISMIR*, Kobe, Japan, 2009

[7] Pikrakis, A., Gmez F., Oramas S. et al.: "Tracking Melodic Patterns in Flamenco Singing by Analyzing Polyphonic Music Recordings" *Proceedings ISMIR*, Porto, 2012.

[8] Serra, X.: "A Multicultural Approach in Music Information Research" *Proceedings ISMIR*, Miami, 2011.

[9] Cornelis, O., Lesaffre, M., Moelants, D., Leman, M.: "Access to ethnic music: Advances and perspectives in content-based music information retrieval" *Signal Processing, 90 (4)*, pp. 1008 -1031, 2010.

Cornelis, O., & Six, J. (2012). Sound to Scale to Sound, a Setup for Microtonal Exploration and Composition. In: Proceedings of the 2012 International Computer Music Conference (ICMC 2012). Ljubljana, Slovenia.

SOUND TO SCALE TO SOUND, A SETUP FOR MICROTONAL EXPLORATION AND COMPOSITION

Olmo Cornelis, Joren Six

Royal Academy of Fine Arts & Royal Conservatory
University College Ghent
Hoogpoort 64, 9000 Ghent - Belgium
olmo.cornelis@hogent.be

ABSTRACT

This paper elaborates on a setup for microtonal exploration, experimentation and composition. Where the initial design of the software Tarsos aimed for the scale analysis of ethnic music recordings, it turned out to deliver a flexible platform for pitch exploration of any kind of music. Scales from ethnic music, but also theoretically designed scales and scales from musical practice, can be analyzed in great detail and can be adapted by a flexible interface with auditory feedback. The output, the scales, are written into the standardized Scala format, and can be used in a MIDI-to-WAV converter that renders a MIDI file into audio tuned in a particular scale. This setup creates an environment for tone scale exploration that can be used for microtonal composition.

1. INTRODUCTION

Pitch and tone scale organization is one of the many interesting musical parameters, but it is foremost the oldest one that has been studied. In ancient Greece, several scientist/philosophers searched for the intimacies of sound and its sympathic vibrations. Using a monochord, Pythagoras revealed the intervallic ratios for his tuning system that was used for ages. It was the beginning of an adventure between mathematical and physical theories versus sounding realities and musicality, where people enigmatically spoke about commas, wolf tones, theory of affects and temperaments. In Western music, music theory developed gradually towards an equal temperament, with some exceptions by experimental composers. In non-Western classical music however, many alternative tuning systems that use specific intervals such as quartertones have been described. In oral music, one encounters even more different scales, that were developed in a master-student relationship, tangled in a functional and societal context and less depending on theoretical framework.

Nowadays music has become a digital object. Individual people, large museums and archives, all of them like to provide digital copies of their analogue recordings to create an easy-to-browse collection. These digital objects also have a completely different research potential. Music and sound can be analyzed as a physical or symbolical string of information; it gave birth to the Music Information Retrieval community, that aimed at automated computational analysis of any musical parameter. In this context, we have developed Tarsos[1]. It was especially designed for the analysis of pitch in African music. The specific characteristics of tone scales in this wide range of music, urged us to develop a very flexible system for pitch analysis, pitch representation, pitch (and scale) interpretation and fitting forms of output. It led to the java platform Tarsos where both automated and manual analysis are implemented in an interface where users can alter any found pitch suggestion, thus listen and verify in order to retrieve the used tone scale. This prior goal soon extended towards small use cases in listening to the unique qualities of retrieved scales. It was a pilot for a more systematic approach to alter Tarsos in such way that it becomes a tool for microtonal experiment as well. Nowadays several options are built in the software that makes microtonal composition much more easy and accessible. Any scale, retrieved from analysis, based on theory, or created from experiment, coupled with a score can be rendered into a WAV-file using a MIDI synthesizer.

This paper is structured as follows: an introduction sketching the background for this research. Chapter two will provide a view on the methodology. Chapter three documents several case studies. The final chapter states aspects on future work.

2. METHODOLOGY

Figure 1 summarizes the setup for this microtonal experiment in a circular triad. The software Tarsos explores and extracts scales, the software Scala gathers all the files and serves as a basis for the auditory feedback, note by note or converting MIDI to WAV-files.

2.1. TARSOS, inner workings

Most software tools for pitch analysis focus on Western music, and thus focus on a pitch organization of 12 pitch classes per octave, in an equal temperament, that is 100 cents per interval. However, to fill the need for pitch analysis of ethnic music, another approach was needed that

[1] Tarsos, together with a manual, the source code and other documentation, can be found at http://tarsos.0110.be.

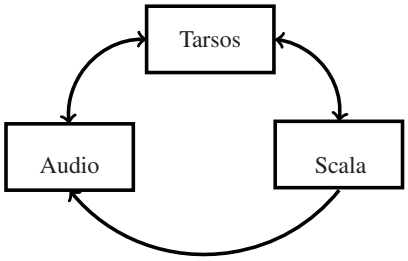


Figure 1. Circular triad between input and output; Tarsos analyses audio, Scala organizes the output of the analyses, which can be uploaded again in Tarsos so any scale can be rebuilt and sonified.

has a much more fine grained range of pitch possibilities. Tarsos has been designed especially for handling microtonal series of pitches. Pitch annotations are extracted from an audio signal by counting how many times each fundamental frequency is repeated throughout that audio signal. Those annotations are processed in such way that they generate musicological meaningful representations, which is not straightforward for ethnic music that relies on completely different musical concept of pitch, pitch perception, scale and its possible meaning. The specific initial intention of Tarsos is to offer a flexible system for pitch annotation by its combination of graphical interface, filtering options, manual and automated annotation processes and direct auditory feedback. The use of a fine-grained histogram that allows up to 1200 values per octave, makes Tarsos convenient for studying pitch deviations in music, and for studying specific tone scales that differ from the equal temperament as appears in many non-Western music. Tarsos has both a graphical user interface, a command line interface and an application programmers interface (API).

For a detailed technical description about Tarsos, see [6]. Summarized: first the audio is analyzed in blocks of 10ms and for each block a fundamental frequency estimation is made by several available pitch trackers namely MAMI[3], YIN[2], MPM[4] and some Vamp plug-ins [1]. Thanks to a modular design, internal and external pitch detectors can be easily added. Secondly, the frequencies are converted to the cents scale with the C below C0 set to zero cents while maintaining a list with the number of times each frequency occurs.

That information is visualized in a number of ways. A first type of visualization is the piano roll representation. In this representation each annotated pitch is plotted over time. A second type of visualization is the pitch histogram, which shows the pitch distribution regardless of time. The pitch histogram is constructed by assigning each pitch annotation in time to a bin between 0 and 14400 cents, spanning 12 octaves. Reducing the annotations to one octave, obtained by adding each bin from the pitch histogram to a corresponding modulo 1200 bin, builds a quasi-continuous pitch class histogram of 1200 values, which is the third type of visualization. Peak detection on the pitch class histogram results in an interval table, that contains the tone scale and all of its intervals. All visualizations and the table interact directly: e.g. changes made to the peak locations propagate instantly throughout the interface.

2.2. SCALA gets the scales

Scala[2] is a software tool for experimentation with musical tunings, such as just intonation scales and non-Western scales. It supports scale creation, editing, comparison, analysis, storage... Exploration of tunings and scales fits Scala very well. The Scala program comes with a dataset of over 3900 theoretical scales ranging from historical harpsichord temperaments over ethnic scales to scales used in contemporary music. Scala files are text files that contain tone scale information. The .scl text file format is defined by the Scala program. Tarsos can parse and export Scala files. Tone scales built within Tarsos that are exported use the Scala standardized .scl format. When used as input for Tarsos, these .scl-files can e.g. be used to compare with histograms that were extracted from audio. Scala files can also be used to compare different tone scales within Tarsos or within the Scala program.

2.3. TARSOS towards sound

There are several options to get sound out of Tarsos. A first one uses the interactive histogram components. One can listen to any histogram in Tarsos by clicking it. A click sends a MIDI-message with a pitch bend to synthesize a sound with a pitch that corresponds to the clicked location within the histogram. A second option is the interval table. In the interval table, one can listen to every detected pitch class or every possible interval between two pitch classes. The interval table, as mentioned before, can be constructed from audio (via peak detection on a pitch class histogram) or from an imported Scala-file.

The third and last option, and perhaps the most practical one, is to use a MIDI keyboard. The MIDI Tuning Standard defines MIDI messages to specify the tuning of MIDI synthesizers. Tarsos can construct Bulk Tuning Dump-messages based scale information - again, either from a Scala-file or from an analzed audio file - to tune a synthesizer. Since most hard- and software synthesizers ignore these messages Tarsos contains the Gervill synthesizer, one of the few software synthesizers that do offer support for tuning messages[3]. This makes it possible to play live on a MIDI keyboard, or to process a MIDI file and render it in an arbitrary tone scale.

[2] Find the website of the Scala software program at `http://www.huygens-fokker.org/scala/`

[3] Another approach to enable users to play in tune with an arbitrary scale is to send pitch bend messages when a key is pressed. Pitch bend is a MIDI-message that tells how much a higher or lower a pitch needs to sound in comparison with a standardized pitch. Virtually all synthesizers support pitch bend but it is hard to imitate a tuned keyboard using only pitch bend messages. Pitch bends operate on MIDI channel level which makes it impossible to play polyphonic music in an arbitrary tone scale on one channel.

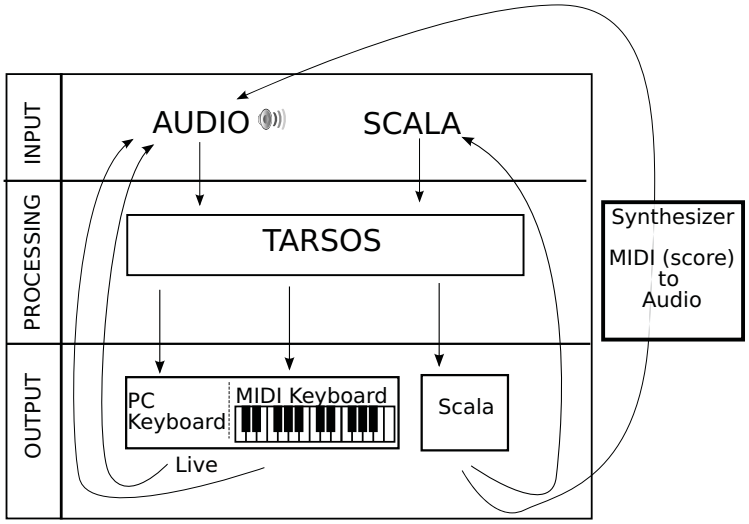


Figure 2. Detailed circular triad as a block diagram for microtonal exploration and composition

Figure 2 shows a conceptual visualization of different interactions and sonifications. As can be seen, it is e.g. possible to start with an audio file audio, export a scala file with a detected tone scale or play a MIDI keyboard.

3. CASE STUDIES

The aim of this research is microtonal composition based on exploration of musical pieces that contain microtonal pitch classes. Therefore some case studies have been chosen to raise a corner of the veil.

3.1. Ethnic scales

Ethnic music offers a unique environment of characteristical timbres, rhythms and textures that need adapted or completely new, innovative tools. The potential of computational research within the context of ethnic music has been stressed by the introduction of the term Computational Ethnomusicology[7]. Hopefully this new interdisciplinary (sub)field can give an impulse to the study and dissemination of a rich heritage of music that is now hidden in archives and aid or even stimulate new musicological field work [6]. As an example for computational pitch analysis, an interesting song is found in the archives of the Royal Museum for Central-Africa (RMCA, Belgium). This song, recorded in Burundi in 1954 by missionary Scohy-Stroobants, is performed by a singing soloist, Leonard Ndengabaganizi. The detected intervals, visualized in Figure 3, are respectively 168, 318, 168, 210, and 336 cents; a pentatonic division that comprises small and large intervals, rather than an equal tempered or meantone division. A capella singing does ensue some variation in pitch classes, but still some particularities can be described: although diverse interval sizes, three nearly fifths are present in the scale. One of these fifths is built by two thirds that resemble a pure minor third and a pure major third (that lies between the intervals $168+210=378$ cents). Thirdly,

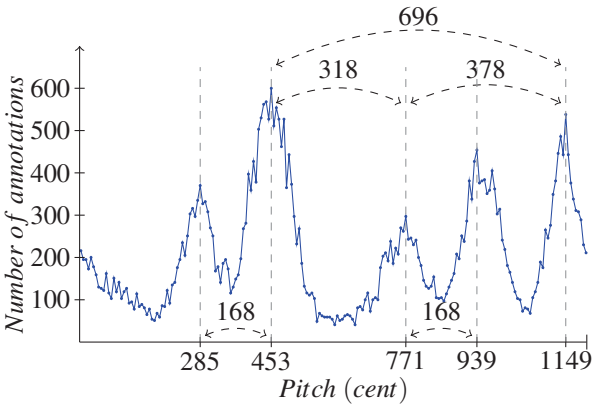


Figure 3. This song uses an unequally divided pentatonic tone scale with mirrored intervals 168-318-168, indicated on the pitch class histogram (horizontal axis visualizes the octave reduced view). Also indicated is a near perfect fifth consisting of a pure minor and pure major third.

an interesting mirrored set of intervals is present: 168-318-168. This phenomena has been encountered several times in the RMCA archive: a scale that is constructed around a set of mirrored intervals. It could be a (unconscious) cognitive process to build a scale around such set of mirrored intervals, but makes it also more convenient to perform for human voice.

3.2. Microtonal composition

Harry Partch, Ivor Darreg, and composers from spectral music really devoted their oeuvre to aspects of microtonality. As a tribute, a composition from Darreg is analyzed here as an example. Tarsos has analyzed and rebuilt the scale that was used in track five from Detwelvulate! *'From beyond the Xenharmonic Frontier'*. This composition uses an equal temperament of 9 tones per octave as

Note	C		D		E		F		G		A		B		C
Pythagorean Tuning		204		204		90		204		204		204		90	
Ptolemaic Tuning		204		182		112		204		182		204		112	
Mean-tone Temperament		193		193		117		193		193		193		117	
Werckmeister I		192		198		108		198		192		204		108	
Equal Temperament		200		200		100		200		200		200		100	

Table 1. Diatonic overview of several historical tuning systems. All interval values are expressed in cents.

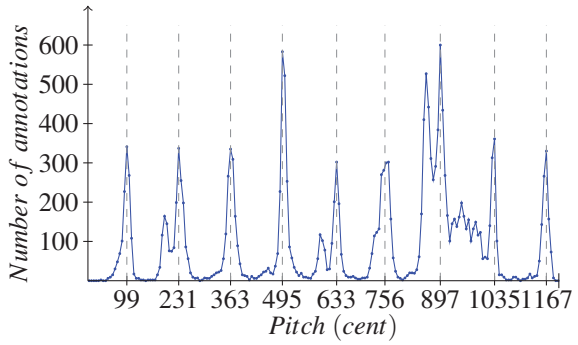


Figure 4. This composition contains an equal temperament of 9 tones per octave

can be seen in Figure 4. Each interval counts 133 cents, which entails the occurrence of 9 major thirds and 9 augmented fifths in the scale as well. It provides the piece a scale that is built on a mixture of unknown and more familiar intervals. However Tarsos did retrieve all nine pitch classes, also three small deviations of pitch classes were noticed. Each of these three pitch classes measure consequently 38 cents lower then the three notes from the intended scale (namely 231 633 and 897 cents). They occurred in a specific octave, and not over the entire ambitus. More research could tell the intention of these tones.

3.3. Historical scales

One can upload any of the historical scales that is listed in Scala, or made manually, and convert any classical symbolic score available in MIDI into audio. As for example Bach's Das Wohltemperierte Klavier, BWV 846-893, can be listened to in equal-temperament or in well-temperament. Interesting opposition since there is still a discussion on which tuning Bach intended these compositions for[5]. Another use case is rendering some baroque compositions, that are known for their sensitivity towards affective theory, in several tuning systems. As a teaser, table1 gives an overview of some historical tunings. Notice the small variations in the different diatonic scales.

4. TARSOS LIVE

Tarsos can be used real-time: when this option is selected, any tone or set of tones that is presented is directly analysized. The scale that is played arises on the graphical axes. By selecting the peaks of the annotations, the program allows you to play together with the live musician in that specific scale. Many possibilities come forward, an interesting one is that Western classical musicians can now play together with any scale that is presented by musicians, ranging from alternative scales to ethnic instruments. Any alteration in the scale is noticed directly, and can the scale can be adjusted.

5. FUTURE WORK

The interface of Tarsos will be provided with a scale visualization that does not refer to the Western keyboard and that comprises the size of the intervals as an ecological user interface. Another feature will be the display of non-octave bound organisation of scales, as for example the 88CET or Bohlen-Pierce. Where the user can (re)set the interval of the octave towards any personal choice. Tarsos will be applied on the entire RMCA archive intending a better insight in African tone scales.

6. REFERENCES

[1] C. Cannam, "The vamp audio analysis plugin api: A programmer's guide," http://vamp-plugins.org/guide.pdf.

[2] A. de Cheveigné and K. Hideki, "Yin, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002.

[3] T. De Mulder, "Recent improvements of an auditory model based front-end for the transcription of vocal queries," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2004.

[4] P. McLeod and G. Wyvill, "A smarter way to find pitch," in *Proceedings of International Computer Music Conference, ICMC*, 2005.

[5] S. M. Ruiz, "Temperament in Bach's Well-Tempered Clavier. A historical survey and a new evaluation according to dissonance theory," Ph.D. dissertation, Universitat Autnoma de Barcelona, 2011.

[6] J. Six and O. Cornelis, "Tarsos - a Platform to Explore Pitch Scales in Non-Western and Western Music," in *Proceedings of the 12th International Symposium on Music Information Retrieval (ISMIR 2011)*, 2011.

[7] G. Tzanetakis, A. Kapur, W. A. Schloss, and M. Wright, "Computational ethnomusicology," *Journal of Interdisciplinary Music Studies*, vol. 1, no. 2, 2007.

Six, J., & Cornelis, O. (2012). A robust audio fingerprinter based on pitch class histograms - applications for ethnic Music archives. In Proceedings of the Folk Music Analysis Conference (FMA 2012), Seville, Spain.

A Robust Audio Fingerprinter Based on Pitch Class Histograms Applications for Ethnic Music Archives

Joren Six*
Royal Academy of Fine Arts & Royal Conservatory
University College Ghent
Hoogpoort 64, 9000 Ghent - Belgium
joren.six@hogent.be

Olmo Cornelis
Royal Academy of Fine Arts & Royal Conservatory
University College Ghent
Hoogpoort 64, 9000 Ghent - Belgium
olmo.cornelis@hogent.be

Abstract

In this paper we present a new acoustic fingerprinting system, based on pitch class histograms. The aim of acoustic fingerprinting is to generate a small representation of an audio signal that can be used to identify identical, or recognize similar, audio snippets in a large audio set. A robust fingerprint generates similar fingerprints for perceptually similar audio signals. A piece of music with a noise added should generate an almost identical fingerprint as the original. The new system, presented here, has some interesting features which makes it a valuable tool to manage ethnic music archives: the fingerprints are rather robust against pitch shift, tempo changes, several synthetic audio effects, and reversal of the audio. To some degree, the system even keeps working when only part of the audio is used to generate the fingerprint.

1 Introduction

In the process of digitizing a large music collection it is possible that the same music is present on different physical media, either as complete copies or as copies of individual tracks. Sometimes it is hard to keep track of which physical media are already digitized and which are still to process. An ability to search for music based on the content of the signal is a valuable tool to prevent duplicates entering the digital version of the music archive. Another use case is to (re)connect meta-data to an audio fragment without any information, but is present in the digital connection.

For large, historical collections of ethnic music the problems sketched above are almost inevitable. Often, individual collections of recordings or discs are donated to museums that lack meta-data. These collections usually are very diverse and several of recordings may already be present in the archive, which is where the need for content based search comes to play. Due to the nature of the original physical media - wax cylinders, wire recordings, magnetic tapes, gramophone records - and the, often abysmal, recording quality a content based search system for ethnic music has to have special features for robustness. Our research is focused on pitch class histograms which appear to be robust enough for the task of acoustic fingerprinting, even in the context of historic ethnic music collections.

This paper is structured as follows: we start with an overview of the system and then argue why it shows potential. Then details about the implementation are unveiled. The third section describes an experiment with the system. The paper ends with a conclusion.

2 System Overview

Figure 1 shows a general acoustical fingerprinting system. Features are extracted from audio and with these features a fingerprint is constructed. The fingerprint is a small representation of the audio. In the best case, perceptually similar audio generates related fingerprints, identical audio should generate identical fingerprints. With the generated fingerprint and a list of previously generated fingerprints an unknown piece of audio is identified. In an ideal system the fingerprints are small but unique for each piece of audio and searching trough a large number of fingerprints is efficient. Alternative systems include the ones described by Haitsma & Kalker (2002); Wang (2003); Allamanche (2001), there is also a review article on audio fingerprinting by Cano et al. (2005).

*Visit http://tarsos.0110.be/tag/fma for more information.

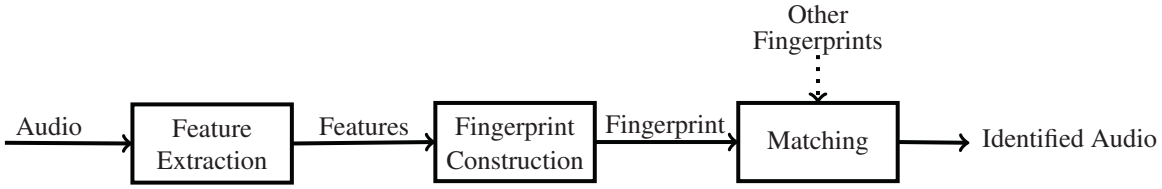


Figure 1: A general fingerprinter.

The workflow of our system, which can be seen in Figure 2, is exactly the same as the general acoustic fingerprinting system but it shows which features are extracted and how a fingerprint is created in our system. The first step is to extract features from audio, in this case a pitch extraction algorithm extracts pitch from audio.

The next step is to create a fingerprint, therefore we use a *pitch class histogram*. A pitch class histogram contains how many times any pitch class has been annotated in a musical segment/piece. A pitch class is defined here as an integer between 0 and 1200, to correspond with the cent unit introduced by Helmholtz & Ellis (1912). If for example, the value of 880Hz has been assigned, this frequency f in Hz can be converted to a cent value c relative to a reference frequency r calculating $c = 1200 \times \log_2(\frac{f}{r})$. With the standard $r = 8.176$Hz[1] this makes $8100 \mod 1200 = 900$ cents. For this block of audio, one value is added to the bin representing 900 cents in the pitch class histogram. If the next block of audio contains for example 220Hz, the exact same thing happens. This is done over and over again for the entire piece.

The third step is to match the constructed fingerprint with a list of previously stored fingerprints. In our system this entails calculating a similarity between probability density functions: the pitch class histograms. For an overview of different ways to do this please consult the overview article by Cha (2007).

As the final step in the process, the identified piece of audio is returned.

2.1 Pitch Class Histograms as Acoustic Fingerprints

There has been a lot of research about pitch class histograms, or very similar concepts under sometimes different names e.g. by Sundberg & Tjernlund (1969); Moelants et al. (2009); Gedik & Bozkurt (2010); Six & Cornelis (2011); Tzanetakis et al. (2002), to name a few. Especially the last article is interesting, in the future work section of they mention the following:

> "*Although mainly designed for genre classification it is possible that features derived from Pitch Histograms might also be applicable to the problem of content-based audio identification or audio fingerprinting (for an example of such a system see Allamanche (2001)). We are planning to explore this possibility in the future.*"

This has, as far as we know, never happened and this article explores this idea. Both Figure 3 and Figure 4 show why this is a reasonable idea. These figures show pitch class histograms of similar but not equal versions of a pentatonic scale. Figure 3 illustrates that pitch class histograms are relatively robust against severe adaptations of the underlying audio: the histogram shape remains more or less the same. Figure 4 shows the result of audio effects which change the pitch. Changing pitch in audio shifts the histogram over the horizontal pitch axis. When calculating a correlation between histograms this needs to be taken into account.

Then histogram overlap or intersection is used as a distance measure because Gedik & Bozkurt (2010) shows that this measure works best for pitch class histogram retrieval tasks. The overlap $c(h_1, h_2)$ between two histograms h_1 and h_2 with K classes is calculated with equation 1. For an overview of alternative correlation measures between probability density functions consult Cha (2007). To calculate the correlation with a pitch shift n equation 2 is used. To make sure that the bin k remains within the bounds of the histogram a $\mod K$ calculation is done. In our application this means that the octave relation is respected, e.g. with n equal to 50 cent[2], the bin at 1170 cent of h_1 is compared with

[1]The MIDI note number standard lets note number 0 correspond with a reference frequency of 8.176Hz, which is C_{-1} with A_4 tuned to 440Hz. If the same reference frequency is used for cents, then MIDI note numbers and cents differ by a factor 100.

[2]Half a semitone, not to be confused with the American rapper.

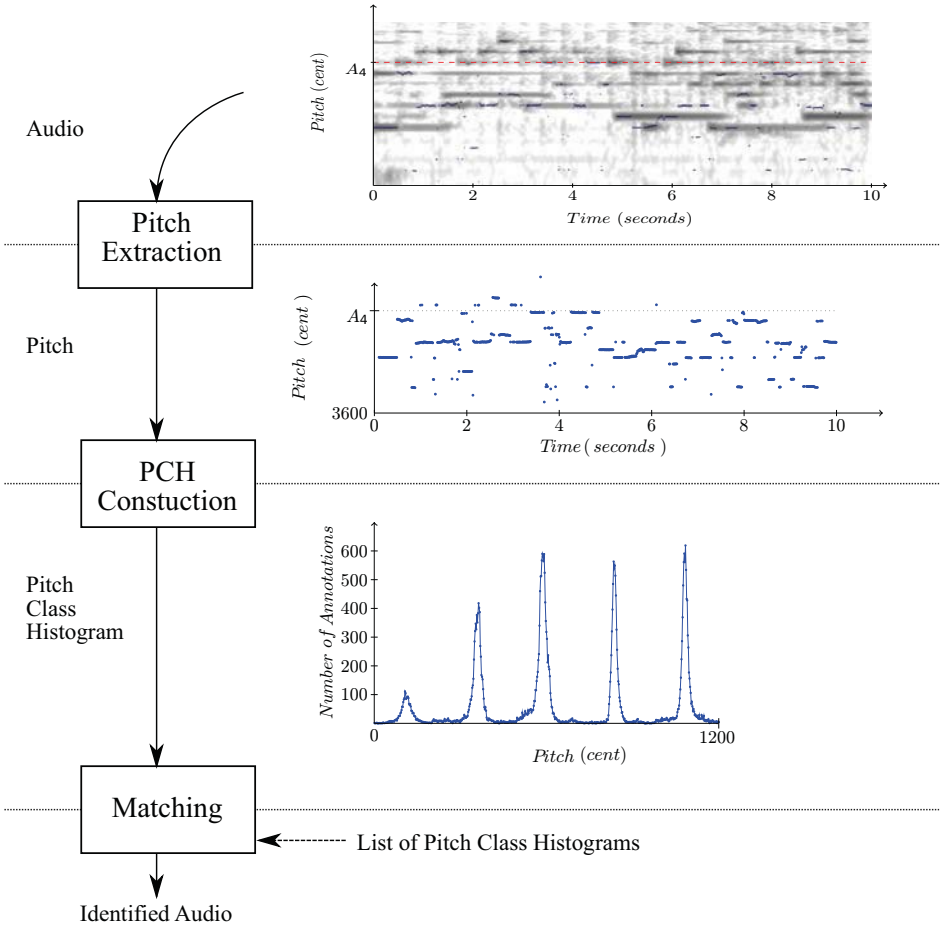


Figure 2: An acoustic fingerprinting scheme based on pitch class histograms.

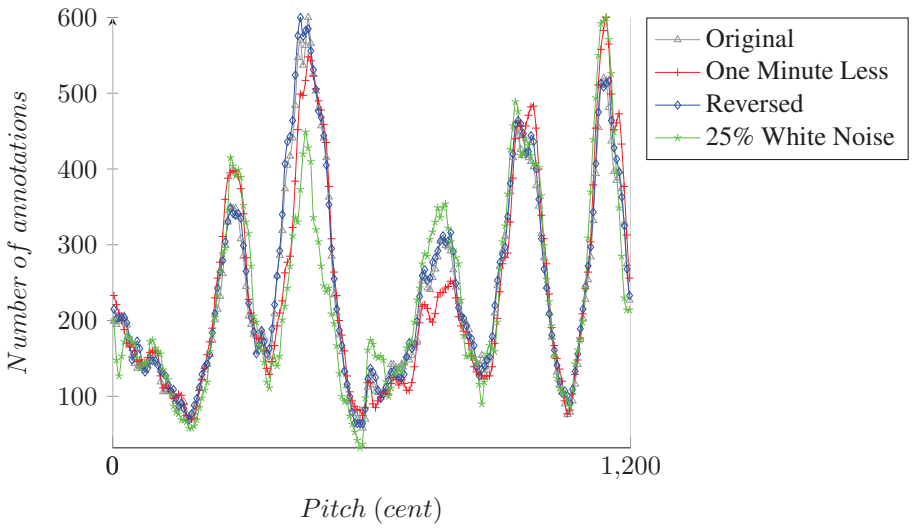


Figure 3: A pitch class histogram of an African song. The histogram of the original song is present, together with a histogram of a reversed, a cropped and noisy rendering of the song. It shows that pitch class histograms are relatively robust against severe mutilations of the underlying audio.

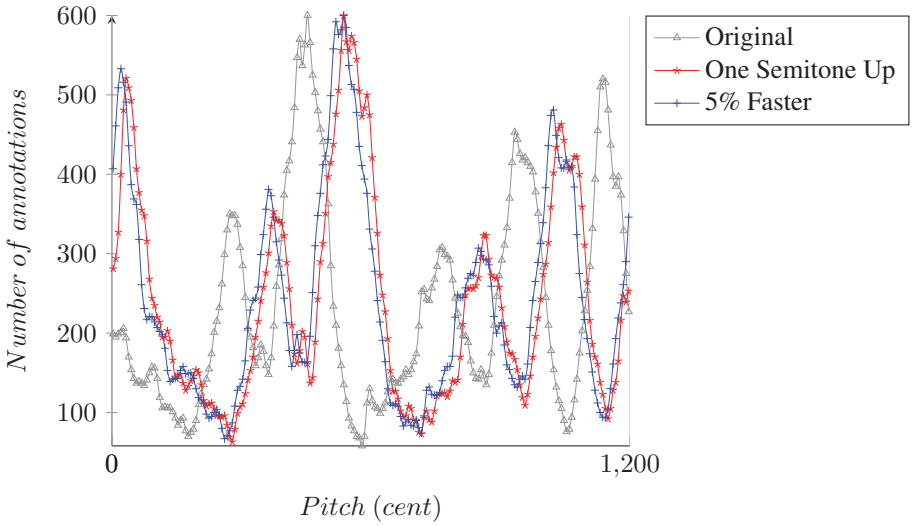


Figure 4: A pitch class histogram of an African song together with a histogram of a version played 5% faster and a pitch shifted version (without affecting the duration). It is clear that almost the same histogram is present three times, only shifted slightly over the horizontal pitch axis.

	5% Faster	One Minute Less	One Semitone Up	Original	Reversed	25% White Noise
5% Faster	1.00	0.92	0.97	0.95	0.97	0.87
One Minute Less	0.92	1.00	0.92	0.94	0.93	0.89
One Semitone Up	0.97	0.92	1.00	0.95	0.96	0.87
Original	**0.95**	**0.94**	**0.95**	**1.00**	**0.96**	**0.89**
Reversed	0.97	0.93	0.96	0.96	1.00	0.88
25% White Noise	0.87	0.89	0.87	0.89	0.88	1.00

Table 1: Similarity between different pitch class histograms of several adapted versions of a song. It shows that the histogram of the song with white noise added differs the most from the original histogram (89%)

the bin at $(1170 + 50) \mod 1200 = 20$ cent of h_2. To find the pitch shift n with maximum correlation, an exhaustive search is done by simply calculating the correlation for each possible shift.

$$c(h_1, h_2) = \frac{\sum_{k=0}^{K-1} min(h_1(k), h_2(k))}{max(\sum_{k=0}^{K-1} h_1(k), \sum_{k=0}^{K-1} h_2(k))} \tag{1}$$

$$c(h_1, h_2) = \frac{\sum_{k=0}^{K-1} min(h_1(k), h_2((k+n) \mod K)}{max(\sum_{k=0}^{K-1} h_1(k), \sum_{k=0}^{K-1} h_2(k))} \tag{2}$$

Table 1 shows the correlation, as defined by equation 2, between the different histograms shown in Figure 3 and 4. It shows that the histogram based on the original version is, for this song, very much alike histogram based on the reversed audio (96%). The version with added noise differs the most from the original (89%). Cropping one minute from the song, which is 7 minutes and 20 seconds long, results in correlation of 94%. The 97% similarity between the 5% faster and pitch shifted version can be explained by the fact that a 5% speed increase translates to a pitch shift of 84 cents which is almost 100 cents. The only difference then is the length of the song, i.e. the number of elements in the histogram, which can be normalized. Section 3 shows if this behaviour is unique to this song or not.

The implementation of the system is done in Java and uses the pitch estimator described in McLeod (2009). For testing purposes, a platform independent version can be downloaded here `http://an.url.left.out.for.review`. There you can also find scripts and data used for this paper.

Modification	First hit	First two	First Three
Original	100%	100%	100%
10% slower	36%	38%	40%
20% slower	0%	2%	6%
10% faster	42%	48%	50%
20% faster	6%	10%	16%
Reversed	100%	100%	100%
One semitone up	96%	96%	96%
One semitone down	86%	92%	94%
Two semitones up	76%	80%	84%
Two semitones down	66%	72%	74%
10s cropped	88%	96%	98%
15s cropped	80%	92%	92%
20s cropped	58%	70%	74%
25s cropped	44%	54%	56%
30s cropped	42%	46%	50%
35s cropped	26%	30%	32%
40s cropped	22%	24%	26%
45s cropped	18%	20%	22%
50s cropped	14%	16%	22%
55s cropped	14%	14%	16%
60s cropped	12%	12%	12%
5% white noise	18%	20%	22%
10% white noise	2%	4%	6%
15% white noise	0%	0%	0%

Table 2: The results for a retrieval task on a data set of 10272 files. 27 effects were applied to 50 songs, generating 1350 modified versions. The goal of the task was to find the original version of the song. The percentages show much of the modified versions were correctly identified in the first, first two, and first three hits. The original and reversed version are retrieved always. The performance on pitch shift and cropping is reasonable, white noise and large tempo changes are problematic.

3 Experimental Results

To show that a fingerprinting scheme based on pitch class histograms has potential, an experiment was done on a data set of 10272 songs from central Africa (see appendix A for more info on the data set). The experiment was constructed as follows: from the data set 50 randomly selected files were copied. A number of modifications and effects - 27 in total - were applied to these 50 files[3], generating 1350 modified songs. The goal of the experiment was correctly match those 1350 songs to the original in the data set.

Table 2 shows the results of the experiment. From those results some conclusions can be drawn. 1) Since the retrieval of the original song always succeeded, it stands to reason that fingerprints for songs are, at least, unique within this data set. An important property of a fingerprint. 2) The reversed audio is also retrieved always, which shows that the pitch estimator used generates almost identical estimations on reversed audio. 3) Pitch shifting works reasonably well. 4) The performance when leaving out the first number of seconds degrades quickly between 15 and 20 seconds. 5) The method does not handle white noise that well. The 20%, 25% and 30% white noise versions were left out of the table since no matches were found.

4 Conclusion & Future Work

In this paper a new approach for acoustic fingerprinting, based on pitch class histograms, was presented. After the introduction, which sketched the applications for the system, an overview of the working principles of acoustic fingerprinting in general and our system in particular were given. The second chapter also explained why pitch class

[3]SoX - Sound Exchange, a command line utility, was used to apply effects to the original file. Following command line instructions were used: `pitch`, `speed`, `reverse` , `trim`, and `synth whitenoise`. For more information on SoX, and the exact meaning of the effects, see `http://sox.sf.net`.

histograms can be used as fingerprints. Some details about the implementation are also given. In chapter three experimental evaluation was done.

This paper has shown that an acoustic fingerprinting system based on pitch class histograms is rather robust and has potential but a lot of questions remain open. The experiment in this paper only discusses a retrieval task for complete songs and for a limited number of effects. Some future work includes:

1. Expand the retrieval task to include more audio (Western music) and apply more audio effects: echo, digital analogue / analogue digital conversions, low bit rate encoding, band pass filtering,...
2. See if the system can be applied to identify small fragments of music instead of complete songs. How small is the minimum fragment? Which adaptations need to be done for broadcast monitoring, processing streams?
3. Experiment with pitch estimators, when the pitch estimator is replaced is there a significant impact on the results?
4. Handle scalability and performance issues. Can the fingerprint size be reduced, without loss of accuracy? Is it possible to speed up the matching step significantly?

As a final remark, we would like to note that this article is rather unique because it presents a generally applicable algorithm that is tested on ethnic music first. Only later it will be applied to western music. This is partly due to the fact that we only have access to a large data set with African music but is also a philosophical statement: instead of adapting techniques used on Western music for applications with ethnic music, why not, for once, do it the other way around?

References

Allamanche, E. (2001). Content-based identification of audio material using mpeg-7 low level description. In *Proceedings of the 2nd international symposium on music information retrieval (ISMIR 2001).*

Cano, P., Batlle, E., Kalker, T., & Haitsma, J. (2005). A review of audio fingerprinting. *The Journal of VLSI Signal Processing*, *41*, 271-284.

Cha, S.-h. (2007). Comprehensive survey on distance / similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, *1*(4), 300–307.

Gedik, A. C. & Bozkurt, B. (2010). Pitch-frequency histogram-based music information retrieval for turkish music. *Signal Processing*, *90*(4), 1049–1063.

Haitsma, J. & Kalker, T. (2002). A highly robust audio fingerprinting system. In *Proceedings of the 3th International Symposium on Music Information Retrieval (ISMIR 2002).*

Helmholtz, H. von & Ellis, A. J. (1912). *On the sensations of tone as a physiological basis for the theory of music* (translated and expanded by Alexander J. Ellis, 2nd English dr.) [Book]. Longmans, Green, London.

McLeod, P. (2009). *Fast, accurate pitch detection tools for music analysis*. Academisch proefschrift, University of Otago. Department of Computer Science.

Moelants, D., Cornelis, O., & Leman, M. (2009). Exploring african tone scales. In *Proceedings of the 10th International Symposium on Music Information Retrieval (ISMIR 2009).*

Six, J. & Cornelis, O. (2011). Tarsos - a Platform to Explore Pitch Scales in Non-Western and Western Music. In *Proceedings of the 12th International Symposium on Music Information Retrieval (ISMIR 2011).*

Sundberg, J. & Tjernlund, P. (1969). Computer measurements of the tone scale in performed music by means of frequency histograms. *STL-QPS*, *10*(2-3), 33-35.

Tzanetakis, G., Ermolinskyi, A., & Cook, P. (2002). Pitch histograms in audio and symbolic music information retrieval. In *Proceedings of the 3th International Symposium on Music Information Retrieval (ISMIR 2002)* (pp. 31–38).

Wang, A. L. (2003). An Industrial-Strength Audio Search Algorithm. In *Proceedings of the 4th International Symposium on Music Information Retrieval (ISMIR 2003)* (pp. 7–13).

A Audio Material

For the comparison of different pitch trackers on pitch class histogram level a subset of the music collection of the *Royal Museum for Central Africa* (RMCA, Tervuren, Belgium) was used. The museum focuses on the African culture

and treasures all kinds of ethnographic objects. The archive of the Department of Ethnomusicology has a digitized collection of about 50.000 sound recordings, with a total of 3000 hours of music, mostly field recordings made in Central Africa of which the oldest going back to 1910. The audio archive is one of the biggest and best documented[4] archives worldwide for the region of Central Africa. A song from the RMCA collection was also used in section 2.1. It has tape number MR.1954.1.18-4 and was recorded in 1954 by missionary Scohy-Stroobants in Burundi.

[4]There is a website about the audio dataset of the Royal Museum for Central Africa featuring complete meta-data and some audio fragments. It can be found at `http://music.africamuseum.be`

Six, J. & Cornelis, O. (2011). Tarsos a Platform to Explore Pitch Scales in Non-Western and Western Music, Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011). Utrecht, The Netherlands.

TARSOS - A PLATFORM TO EXPLORE PITCH SCALES IN NON-WESTERN AND WESTERN MUSIC

Joren Six and Olmo Cornelis
Royal Academy of Fine Arts & Royal Conservatory,
University College Ghent
joren.six@hogent.be - olmo.cornelis@hogent.be

ABSTRACT

This paper presents Tarsos [1] , a modular software platform to extract and analyze pitch and scale organization in music, especially geared towards the analysis of non-Western music. Tarsos aims to be a user-friendly, graphical tool to explore tone scales and pitch organization in music of the world. With Tarsos pitch annotations are extracted from an audio signal that are then processed to form musicologically meaningful representations. These representations cover more than the typical Western 12 pitch classes, since a fine-grained resolution of 1200 cents is used. Both scales with and without octave equivalence can be displayed graphically. The Tarsos API [2] creates opportunities to analyse large sets of - ethnic - music automatically. The graphical user interface can be used for detailed, manually adjusted analysis of specific songs. Several output modalities make Tarsos an interesting tool for musicological analysis, educational purposes and even for artistic productions.

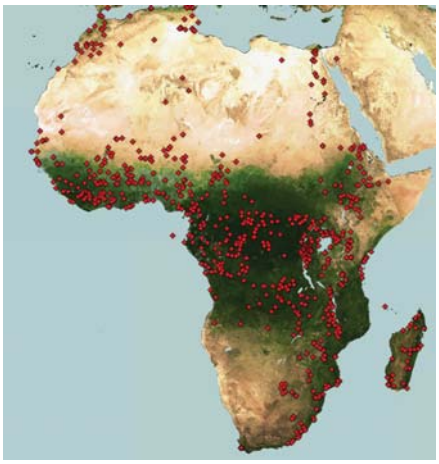


Figure 1. A visualization of the music archive of the Royal Museum for Central Africa. The dots mark places where recordings have been made.

1. INTRODUCTION

A 2007 f(MIR) article by Cornelis et al. [3] argued that access to ethnic music could become one of the next big challenges for the MIR community. It gives an overview of the difficulties of working with ethnic music: i) There is an enormous variety of styles, timbres, moods, instruments falling under 'ethnic music' umbrella. ii) The absence of a theoretical framework and a different attitude towards music imply that western music-theory concepts do not always apply. iii) A third factor that complicates access to ethnic music is its distribution. Archives of ethnic music are often not or not yet digitized, badly documented, and when metadata is available terminology and spelling may vary. These three elements cause a lack of ground truth, which makes (large scale) research challenging.

There are difficulties working with ethnic music but there is also a lot of potential. While some archives with ethnic music are being digitized, the need for specialized MIR applications becomes more apparent. Ethnic music offers a unique environment of characteristic timbres, rhythms and textures which could use adapted or completely new, innovative tools. The potential of computational research within the context of ethnic music has been stressed by the introduction of the term *Computational Ethnomusicology* [13]. Hopefully this new interdisciplinary (sub)field can give an impulse to the study and dissemination of a rich heritage of music that is now hidden in archives and aid or even stimulate new musicological field work.

This research focuses on one of those unique collections of ethnic music: the audio archive of the Royal Museum for

[1] Tarsos is open source and available on http://tarsos.0110.be. It runs on any recent Java Runtime and can be started using Java Web Start.

[2] With the Application Programmers Interface tasks can be automated by programming scripts. For an application see chapter 5.

Central Africa (RMCA) in Belgium. It is one of the largest collections worldwide of music from mainly Central Africa. Figure 1 displays the geographical distribution of the audio collection [3] that consists of about 50,000 sound recordings (with a total of 3,000 hours of music), dating from the early 20th century up to now. A selection of this data set with African music has already been used for a study on pitch organization and tone scales [9].

This paper is structured as follows: After this introduction sketching the background for this research the following chapter identifies the need for precise pitch analysis and the rationale behind the development of Tarsos. Chapter three will provide a view on the method we use, and give a brief overview of related work. Chapter four documents the structure and method of the Tarsos platform. Example applications of Tarsos can be found in chapter five. The final chapter gives a conclusion and ponders on future work.

2. SCALE ORGANISATION

For *Western music* pitch organization in music relies on a well-defined, historically grown music theory. Nowadays almost all western music relies on a division of the octave in 12 equal parts. Only few composers have explored different divisions of the octave (e.g. Darreg, Partch).

In *non-Western classical music*, tone scale organization leans on an, often very different, theoretical system than the Western equal temperament. The most outspoken difference is that not all pitch intervals have an equal size. This can result in an explicitly sought musical tension. An example is the unequal octave division of the Indonesian gamelan Pelog scale.

Oral music traditions however, rely almost exclusively on musical practice. Without a written musical theory the master-student relationship becomes very important, together with the societal context in which people hear music. An oral culture does not support the development towards a polished music theory but grows more organically. These factors define the specific characteristics of the music itself: less harmonic impact, instruments with varying tuning, no harmonic modulation and a large number of different tone scales. Until now, far too little attention has been paid to this tone scale diversity. There is a need for a system that can extract pitch organisation - scales - from music in a culture-independent manner.

Currently there is software available for pitch analysis but it mainly focuses on Western music and is used for e.g. key-detection in pop music. To fill the need for automated pitch analysis of ethnic music *Tarsos* has been developed. Tarsos creates opportunities to analyse pitch organization in large music archives, document tone scales and find patterns in pitch usage.

[3] There is a website featuring complete descriptions and audio fragments, it can be found at `http://music.africamuseum.be`

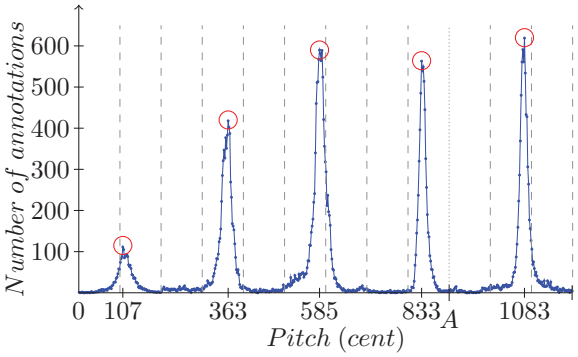


Figure 2. A pitch class histogram that shows how much pitch classes are present in a piece of music. The graph shows absolute pitch annotations collapsed to one octave. The circles mark the most used pitch classes. For reference, the dashed lines represent the Western equal temperament. The pitch class A is marked with a dotted line.

3. PITCH ANALYSIS

The basic idea behind the method we use is simple: count how many times each fundamental frequency is repeated throughout an audio signal and represent this data in a useful way. This method has a long tradition and historically this was done by hand, or, more anatomically correct, by ear. Each tone in a musical piece was compared with a large set of tuned bells and every match was tallied. This method is very labour-intensive and does not scale to large music archives.

Already in the late sixties researchers automated this process to study the tone scale of a Swedish folk instrument [11]. Since then various terms have been introduced to describe this, or closely related ideas: Frequency Histogram [11], Chromavector [9], Constant-Q Profile [10], Harmonic Pitch Class Profile [5] and Pitch-frequency Histogram [6].

Working with ethnic music, and especially African music, it is important that the pitch organization diversity can be captured. In [9] this is done as follows. At first the audio is analysed in blocks of 10ms and for each block a fundamental frequency estimation is made. Secondly, the frequencies are converted to the cents scale with C_0 set to zero cents while maintaining a list with the number of times each frequency occurs. And finally the listed values are reduced to one octave. This results in a quasi-continuous *pitch class histogram* of 1200 values as seen in Figure 2. With Tarsos this method is automated in a flexible way.

Pitch class histograms can be used for various applications. The most straightforward application is tone scale

detection. To extract a scale from a pitch class histogram peak extraction is used: i.e. finding the circles in Figure 2. With the pitch classes identified a pitch interval matrix can be constructed and subsequently used for comparison and analysis.

4. TARSOS PLATFORM

The main contribution of this paper is Tarsos: a platform for pitch analysis. It makes the methods described in [6, 9] easier to use and therefore accessible to a larger audience. Essentially Tarsos tries to make one-off studies of pitch usage easily repeatable, verifiable and scalable to large data sets. The functions of Tarsos will be explained using the block diagram in Figure 3. This should make the information flow clear and provide a feel on how Tarsos can be used.

4.1 Input

As input Tarsos accepts audio in almost any format. All audio is transcoded to a standardized format. The conversion is done using FFMPEG [4], the default format is PCM WAV with all channels are downmixed to mono.

Another input modality are Scala files. Scala files are standardized text files which contain tone scale information. The file format is defined by the Scala program. Quoting the Scala website: `http://www.huygens-fokker.org/scala/`

> "Scala is a powerful software tool for experimentation with musical tunings, such as just intonation scales and non-Western scales. It supports scale creation, editing, comparison, analysis, storage, ... Scala is ideal for the exploration of tunings and becoming familiar with the concepts involved."

The Scala program comes with a dataset of over 3900 scales ranging from historical harpsichord temperaments over ethnic scales to scales used in contemporary music. Tarsos can parse and export scala files. Their use should become apparent in section 4.5.

4.2 Analysis

During analysis each block of audio is examined and zero, one or more fundamental frequencies are assigned. The block size and the number of extracted frequencies depend on the underlying fundamental frequency detection algorithm. Several detection algorithm implementations are distributed together with Tarsos and thanks to its modular design new ones can be added. For practical purposes platform-independent - pure Java - implementations of YIN [4] and MPM [8] are available without any configuration. Currently there is also support for the MAMI-detector [2] and for any VAMP-plugin [1] that generates frequency annotations. These external detectors are platform dependant and need some configuration but once correctly configured their use is completely transparent: the generated annotations are transformed to a unified format, cached and then used for representation.

[4] FFmpeg is a complete, cross-platform solution to record, convert audio and video. It has decoding support for a plethora of audio formats.

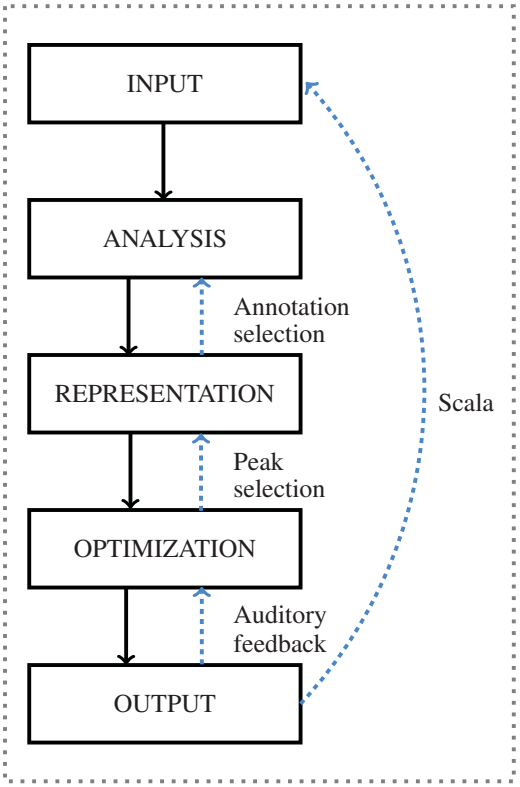


Figure 3. The main flow of information within Tarsos.

4.3 Representation

The most straightforward representation of pitch annotations is plotting them over time. This results in a *piano-roll* like view. In monophonic music this visualizes the melody. In polyphonic music it shows information about harmonic structures and the melodic contour. The piano-roll aids transcription and makes repeating melodic patterns visible. Figure 4 shows a screenshot of Tarsos, the piano roll representation is marked with 3. With the interactive user interface the piano roll representation can be used to select an area of annotations you are interested in. This can be used to ignore annotations below a certain pitch threshold (e.g. pitched percussion) or to compare the first part of a song with the second part. The selection - represented by the upwards arrow between analysis and representation in Figure 3 - influences the next representation.

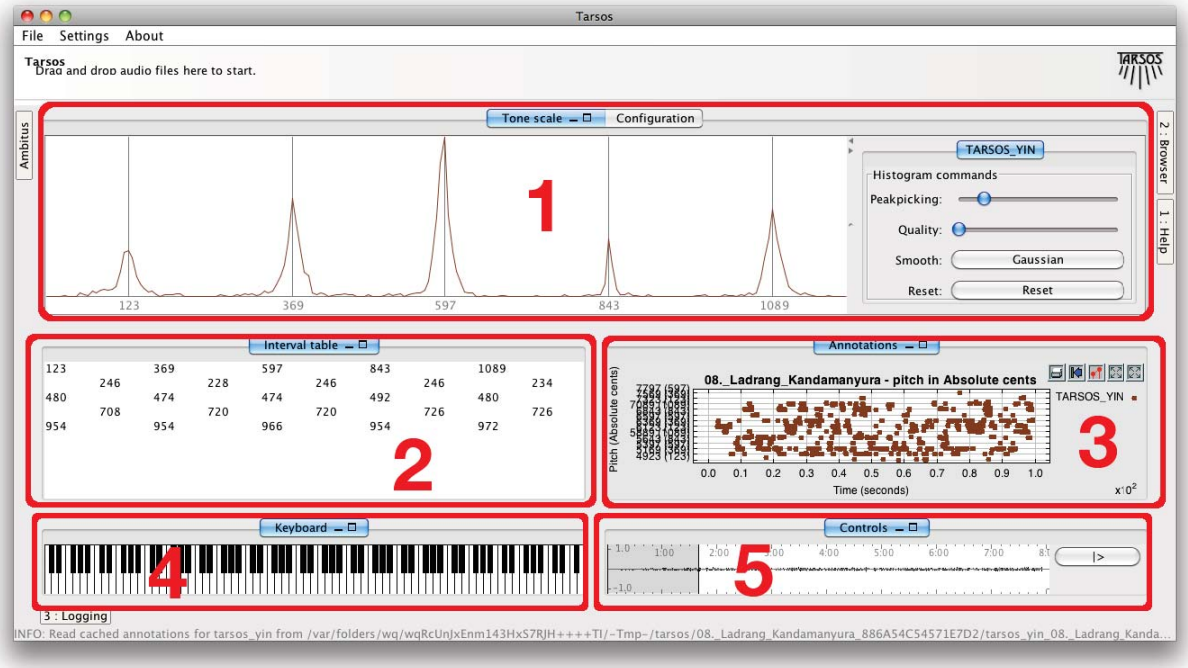


Figure 4. A screenshot of Tarsos: 1) a pitch class histogram, 2) a pitch class interval table, 3) a piano roll like view on annotations, 4) a MIDI keyboard and 5) a waveform. Tarsos is available on `http://tarsos.0110.be`.

Within Tarsos the *pitch histogram* is constructed by assigning each annotation to a bin between 0 and $12 \times 1200 = 14400$ cents, spanning 12 octaves. The height of each peak represents the total duration of a particular detected absolute pitch within a piece. As mentioned in section 3 to transform the pitch histogram to a *pitch class histogram* all values are folded to one octave. In the pitch class histogram a peak represents the total duration of a detected pitch class within a piece. An example of a pitch class histogram can be seen in Figure 2 or the area marked with 1 in Figure 4.

A more high level, musicologically more meaningful representation is the *pitch interval matrix*. It is constructed by applying automatic or manually adjusted peak detection on the pitch class histogram and extracting the positions of the pitch classes. It contains the tone scale of a song and the intervals between the pitch classes.An example of a pitch interval matrix extracted from the pitch class histogram in Figure 2 can be seen in Table 1. In the screenshot, Figure 4 it is marked as 2.

4.4 Optimisation

Automatic peak extraction may yield unwanted results. Therefore there is a possibility to adjust this process manually. Adding, removing or shifting peak locations is possible with the pitch class histogram user interface. Changing the position of a peak has an immediate effect on all other representations: the pitch interval matrix is reconstructed, the reference lines in the pitch histogram and piano roll are adjusted accordingly.

4.5 Output

Tarsos contains export capabilities for each representation, from the pitch annotations to the pitch class interval matrix there are built-in functions to export the data, either as comma separated text files or as image files. Since Tarsos has a scriptable, documented API which can be accessed by any Java Virtual Machine (JVM) compatible programming language - Groovy, Scala [5], Clojure, Java - there is also a possibility to add new output formats based on the internal object model. Scripting is also the way to go when processing a large number of files.

As previously mentioned, for pitch class data there is a special standardized text file format defined by the Scala program: the scala file with the `.scl` extension. Scala files can be used to compare different tone scales within Tarsos or with the Scala program. When used as input for Tarsos, these files provide a reference for the pitch class histogram extracted from audio. A scala file e.g. extracted from Figure 2 with pitch classes $(107, 363, 585, 833, 1083)$

[5] Do not confuse the Scala programming language with the Scala software tool for scale analysis. Information about the programming language can be found at `http://scala-lang.org`

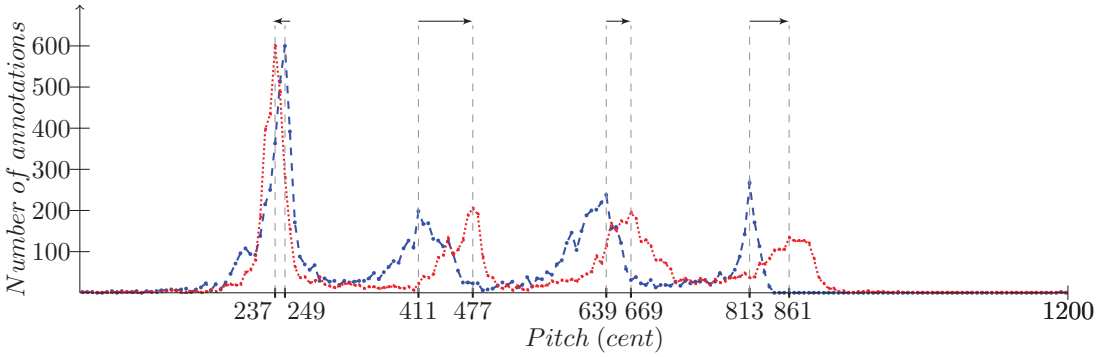


Figure 5. MR.1973.9.41-4, the second minute of the song is represented by the blue, dashed line, the seventh by the red, dotted line. Comparing the second with the seventh minute shows that during the performance the fiddle player changed hand position. The lowest, most stable pitch class is the result of the open string which lost tension during the piece and started to sound lower, in stark contrast with the other pitch classes.

Pitch Class (cent)	Interval (cent)
107	
	255
363	478
	222 725
585	470 976
	248 720
833	498
	251
1083	

Table 1. A pitch interval matrix with pitch classes and pitch class intervals, both in cents. The peaks detected in Figure 2 are used.

can be used to compare pitch use of a song recorded in the same geographical area: do they both use the same absolute pitch or the same pitch intervals?

The pitch annotations can also be synthesized. This results in an audio file which can be used to check if the annotations make sense. Overlapping the original sound with the synthesized annotations makes clear when no, or incorrect annotations were made and, conversely when annotations are correct. This auditory feedback can be used to decide if the annotations are trustworthy (the upwards arrow starting from output in Figure 3).

A completely different output modality is MIDI. The MIDI Tuning Standard defines MIDI messages to specify the tuning of MIDI synthesizers. Tarsos can construct `Bulk Tuning Dump`-messages based on extracted pitch class data to tune a synthesizer enabling the user to play along with a song in tune. Tarsos contains the Gervill synthesizer, one of the few (software) synthesizers that offer support for tuning messages.

5. APPLICATIONS

This section illustrates how Tarsos enables or facilitates research on pitch use. The examples given could inspire third party users - musicologists - to try Tarsos and use it to solve their own research questions.

A first example is an analysis on a single African fiddle piece. In African music pentatonic scales are common but this piece uses a tetratonic scale as seen in Figure 5. The scale is a result of a playing style with three - more or less equally spaced - fingers and an open string. The graphical interface of Tarsos was used to compare the second minute of the song with the seventh, this can be accomplished by selecting the annotations in the piano roll window. This shows that the open string lost tension during the performance - it started to sound lower - in stark contrast with the other pitch classes. The results were exported using the LaTeX-export function and are shown in Figure 5.

A second example illustrates what can be done with a script that processes a lot of audio files in batch and the Tarsos API. In an article by Bozkurt [6] pitch histograms are used for - amongst other tasks - makam [6] recognition. The task is to identify which of nine makams is used in a specific song. A simplified, generalized implementation of this task was scripted. With this script it is possible to correctly identify 39% of the makams using a dataset of 800 files. Some makams look very much alike: if the first three guesses are evaluated the correct makam is present in 75% of the cases. The example is fully documented in the Tarsos manual available on the website `http://tarsos.0110.be`, also the source code is available there. This method is very general and directly applicable to e.g. harpsicord tuning estimation as done, using another approach, by Tidhar et al [12].

[6] A maqam defines rules for a composition or performance of classical Turkish music. It specifies melodic shapes and pitch intervals.

6. CONCLUSION, DISCUSSION AND FUTURE WORK

In this paper Tarsos was presented, a modular software platform to extract and analyze pitch organization in music. After an introduction explaining the background and the needs for precise pitch analysis, chapter two provided some context about the method used and points to related work. Chapter three gave a high level overview of the different components of Tarsos.

Currently Tarsos offers a decent foundation for research on pitch but it also creates opportunities for future work. One research idea is to reintroduce time domain information. By creating pitch class histograms for a sliding time-window and comparing those with each other it should be possible to detect sudden changes in pitch usage: modulations. Using this technique it should also be possible to detect and document pitch drift in choral or other music on a large scale. Automatic separation of speech and music could be another application.

Another research area is to extract features on a large data set and use the pitch class histogram or interval data as a basis for pattern recognition and cluster analysis. Using Tarsos' scripting abilities with a timestamped and geotagged musical archive it could be possible to detect geographical or chronological clusters of similar tone scale use.

On the longer term we plan to add comparable representations of other musical parameters to Tarsos as well. In order to compare rhythmic and instrumental information, temporal and timbral features will be included. Our ultimate goal is to develop an objective, albeit partial, view on music by combining those three parameters.

During this type of research one should keep this quote in mind:

> "Audio alone might not be sufficient to understand ethnic music. What does it mean to describe music from a culture where the word "music" exists only in connection to body movement, smell, taste, colour. The idea of separating sound from the rest of its physical environment (movement, smell, taste, colour) may well be a weird "invention" of the West. We cannot understand ethnic music correctly without its social function and context [7]."

However, we do can gain interesting insights and alleviate accessibility problems, which is what we are aiming for.

7. REFERENCES

[1] Chirs Cannam. The vamp audio analysis plugin api: A programmer's guide. `http://vamp-plugins.org/guide.pdf`.

[2] L. P. Clarisse, J. P. Martens, M. Lesaffre, B. De Baets, H. De Meyer, and M. Leman. An auditory model based transcriber of singing sequences. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 116–123, 2002.

[3] Olmo Cornelis, Dirk Moelants, and Marc Leman. Global access to ethnic music: the next big challenge? In *Proceedings of 9th ISMIR Conference*, 2009.

[4] Alain de Cheveigné and Kawahara Hideki. Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930, 2002.

[5] Takuya Fujishima. Realtime chord recognition of musical sound: A system using common lisp music. In *Proc. Int. Comput. Music Conf*, pages 464–467, 1999.

[6] Ali C. Gedik and Barış Bozkurt. Pitch-frequency histogram-based music information retrieval for turkish music. *Signal Processing*, 90(4):1049–1063, 2010.

[7] Micheline Lesaffre, Olmo Cornelis, Dirk Moelants, and Marc Leman. Integration of music information retrieval techniques into the practice of ethnic music collections. In *Proceedings Unlocking Audio 2*, 2009.

[8] Phillip McLeod and Geoff Wyvill. A smarter way to find pitch. In *Proceedings of International Computer Music Conference, ICMC*, 2005.

[9] Dirk Moelants, Olmo Cornelis, and Marc Leman. Exploring african tone scales. In *Proceedings of 9th ISMIR Conference*, 2009.

[10] Hendrik Purwins, Benjamin Blankertz, and Klaus Obermayer. Constant Q profiles for tracking modulations in audio data. In *International Computer Music Conference*, pages 407–410, 2001.

[11] J. Sundberg and P. Tjernlund. Computer measurements of the tone scale in performed music by means of frequency histograms. *STL-QPS*, 10(2-3):33–35, 1969.

[12] Dan Tidhar, Matthias Mauch, and Simon Dixon. High precision frequency estimation for harpsichord tuning classification. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 61 –64, march 2010.

[13] G. Tzanetakis, A. Kapur, W. A. Schloss, and M. Wright. Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*, 1(2), 2007.

Cornelis O., Moelants D., Leman M. (2009), Global access to ethnic Music: the next big challenge? Proceedings of the 10th ISMIR, Kobe, Japan.

GLOBAL ACCESS TO ETHNIC MUSIC: THE NEXT BIG CHALLENGE?

Olmo Cornelis
University College Ghent
Olmo.Cornelis@hogent.be

Dirk Moelants
Ghent University
Dirk.Moelants@UGent.be

Marc Leman
Ghent University
Marc.Leman@UGent.be

ABSTRACT

Although MIR is a relatively new branch of research, it has already approached several musical parameters and many musical styles, using different methodologies, from low-level signal analysis up to higher-level symbolic and semantic approaches.
Several implementations have come forward, but when we want to apply these tools to ethnic music we are often confronted with fundamental problems. This music mostly does not rely on the common Western musical concepts. Therefore most MIR-applications need to be redesigned in order to give a relevant contribution to the analysis and classification of non-Western music.
In an interdisciplinary approach, engineers and ethnomusicologists should be able to achieve considerable progress in the approach of music that lacks commercial means and is difficult to access for the general public. Developers can find new challenges, while the development of interesting tools can give a new impulse to the study and dissemination of a rich heritage of music now hidden in archives.

1. INTRODUCTION

The systematic study of ethnic music has started in the late 19th century. From then on, field work has been conducted all over the world, and numerous sound recordings have been made and documented. However, retrieval and access of ethnic music audio documents is an area that did not get a lot of attention until recently. Worldwide, thousands of cultures have developed their own musical culture, with specific qualities and specific purposes. This makes the field of ethnic music very broad, with a whole range of different timbres, moods, styles, instruments and musical characteristics. Different cultures also develop a different attitude towards music as a phenomenon, which implies that western musical concepts do not always fit ethnic music, thus in many cases music as a concept is not separable from e.g. movements (dance) or texts. Yet, a written conceptual framework of the music is mostly not present. Some 'classical' traditions, as found e.g. in India or the Middle-East, do have a music theory that allows classification according to established concepts (though still fundamentally different from western concepts), but in many cultures music exists without any common theoretical ground. So, mostly, the musical system is not as well-defined as it is for Western music, which makes it hard to set out general rules for analysis.

Besides the enormous variety and the lack of a theoretical framework, a third element that complicates the access to ethnic music is its distribution. A lot of musical cultures are under strong pressure of Westernization; the traditional music thus became a fragile cultural heritage. It is not produced in a commercial circuit in which labelling and classification are omnipresent. Rather, ethnic music is collected by researchers and gathered in archives, which are often not digitized and thus not easily accessible. Sometimes entire collections are really badly documented, and even if the documentation is fully available, the metadata fields are focused on different descriptive aspects than is the case for Western music. Moreover, spelling and terminology may vary. All these cause a lack of coherent ground truth, which makes larger scale research very difficult.

Before going further into the problems and opportunities in the application of MIR techniques in ethnic music, we should define more exactly what we mean with ethnic music, as it is does not point to a specific musical style, but is a term used to group a myriad of different styles and genres.

2. WHAT IS ETHNIC MUSIC?

As ethnic music is not a category that exists as such, but rather a kind of meta-category under which music with completely different origins, functions and musical characteristics is grouped, it is difficult to give a precise definition. Moreover, in common practice there is often a confusion between ethnic music in strict sense, which comprises the music of cultures without written tradition and the broader subject field of ethnomusicology which commonly also includes non-Western classical music, the religious and court music of cultures with a written musical culture, such as China, India or the Middle-East, and folk music, which can be defined as music from cultures with a written musical culture that does not belong

to the classical tradition. The latter category is not restricted to Western folk music, but e.g. also includes Japanese folk songs. Here we see another common point of confusion, namely that between ethnic and non-Western. Some Western music can be called ethnic (at least if we want to include folk music), while some non-Western music is modeled to Western popular or classical music. Sometimes it is difficult to draw a line, as traditional style characteristics often show up, both in works of non-Western classical composers and in local styles of popular music. From the viewpoint of MIR, the main distinction would be in whether or not the music is produced for distribution outside the immediate community. Therefore we exclude any form of notated music as well as commercial popular music, even if it concerns non-Western (e.g. Chinese or Brazilian) popular music or Westernized styles of 'world music'.

3. ACCESS TO ETHNIC MUSIC

3.1. Collection and storage of ethnic music

Following our definition, ethnic music is produced for use within the community and not for a broader distribution. Scores and recordings are originally not part of the musical culture, but are made by visiting ethnomusicologists. These recordings are not primarily meant for distribution, rather for preservation and research. Consequently, they are stored in archives and historical collections, mostly on a mix of different types of carriers. In recent years efforts have been made to start up digitization projects, but this did not always improve the accessibility, as the files are still part of (closed) collections of institutes or musea. Popular music has a strong link with commercial, industrial activities, steering mutual interaction with technological innovation and societal development. Ethnic music is considered as a museum artefact, a silent memory of an oral culture. It thus needs governmental financial support, and it strikes that (governmental) institutions from different countries tend to follow different policies for presentation, access and storage which makes it even harder to develop fitted research tools.

3.2. The access path itself

Access path to music is the path that exists between a search idea and the delivery of an audio file, with its associated contextual information. The search and retrieval of ethnic music requires strategies that deal with a large variability of music, users, search intentions and expectations. The music's metadata can be fundamentally different from Western standards, and the individual users can have very different backgrounds and have very different intentions in searching certain music. We could distinguish three groups [1]: (1) people from the general public with an interest in ethnic music, but without a very elaborate background. These users typically want to retrieve music using a rather vague and general labelling, such as 'drumming', 'trance music' or 'some song from Rwanda'. A second group consists of users from within the culture. They may have a good knowledge of certain repertoires and functions of the music, and therefore, they tend to ask very specific questions such as: music played by a specific performer, music from one particular village, lyrics, genres, instruments (in local terminology). Finally, the third group of users consists of researchers, who use the database for further study. This group would typically tend to ask questions related to the geographical spread of certain instrument types, or the relative importance of certain rhythmic or pitch musical structures in different regions.

4. MIR AND ETHNIC MUSIC: A COLORFUL BLEND?

Currently, Music Information Retrieval tools have mainly been developed for Western popular and classical music. Ethnic music does however have a link with MIR. Some aspects of the traditional ethnomusicology can claim some commonalities with MIR: audio (field) recordings take a central place in ethnomusicology, leading to collections of audio recordings with their associated meta-data. Ethnomusicologist always had to develop methods to analyze and transcribe these recordings, aiming at a comparison and classification of the music.

The potential of computational research within the context of ethnic music has been stressed by the introduction of the term 'Computational Ethnomusicology' [2]. Ethnomusicology is by nature an interdisciplinary research field, in which diverse fields like musicology, anthropology and acoustics come together. The existence of large audio collections invites to develop automated approaches and its complex contents urges for the development of domain specific techniques to approach the particular characteristics and constraints of different ethnic music styles.

5. MIR POTENTIAL

Music Information Retrieval offers a lot of new potential for researching ethnic music. The challenges can be divided into three very different areas, that can act as methodology: i) low-level based signal processing, ii) high-level research on interpreted and structured audio-data, iii) metadata based retrieval.

- Low-level signal processing extracts information directly from the audio signal. This can give us characteristics of the music, without the need for a theoretical framework, which is very helpful for approaching music of which the structural characteristics are not well-defined. Eventually, this can lead to a query-by-example system, allowing to retrieve music which is largely unfamiliar to the users. Some papers published within the ISMIR community do work on this low-level signal processing [3][4][5].

- The second type of research would focus more onto the interpretation of the music. Within this methodology two paradigms can be handled: A model-driven paradigm starts from analysing specific musical parameters, refining the tools towards the model put forward. This type of model can be used on smaller sets of audio in order to obtain a very detailed description of the sets, but can not easily be transferred to another collection of music, with very different characteristics. On the other hand, a 'serendipity-paradigm' aims to analyze and optimize the algorithms without any steering theoretical framework, but relying on the potential of the merging of tendencies, patterns and similarities when working on large sets of audio. This type of approach allows more flexibility and the treatment of heterogeneous materials, but can not lead to very specific classifications. Papers depending on the model-driven paradigm [6][7] and on the serendipity paradigm [8][9] can be found.
- The third area focuses on flexible data querying. Current tools for query-by-data can be used but need to be extended because the meta-data of ethnic music focuses on other aspects. For example, it is usually not relevant for ethnic music to search for composers and genres, while it is of utmost importance to search for geographical information and social relevance (function). Another remarkable aspect is the lack of standardisation of spelling in the ethnic context, needing a flexible searching tool, using fuzzy logic techniques. A few papers within ISMIR deal on the description of meta-data [10], but no efforts within the field of flexible data querying.

6. MUSICAL PARAMETERS

Pitch organisation reveals the most obvious example of the shortcoming of Western music theory and notation in the field of ethnic music. In ethnic music, the 100 cents based scale is not an absolute factor, and other ways to describe pitch intervals are necessary to grasp the through pitch organization of the music. Some interesting work can be found on actual used pitch scales [1][6]. Aside of the problematic of pitch distribution, in some cultures even the notion of discrete pitch categories is not appropriate. In some cases, a continuous pitch organisation, based on sliding pitches, leading to an organisation in prototypical musical gestures is more adequate [11][12].
Timbre is a musical parameter that can be investigated by low-level signal processing, but can also fit within the serendipity paradigm as described above. Fact is that timbre doesn't have a strict underlying theoretical framework. Semantically, only few words can be connected with timbre aspects. Mostly such matter is i) described in metaphorical terms or ii) displayed in the form of self-organising maps [3][9].
Despite well-organised underlying theoretical structures, the phenomenon of temporal perception can be ambiguous even in a Western perspective [13]. In some cases the analysis of the movements associated with the music, gestures, spoken word and dance come together to form the correct rhythmic interpretation [14] [15]. Summarising, reducing concepts and characteristics of ethnic music onto familiar Western musical parameters may loose much of the rich, dynamic and lively texture of this music because the different underlying structural development.

7. FUTURE WORK

Although ethnic music did not get so much attention in the Music Information Retrieval community, the existence of large archives, with complicated and often deficient meta-data, provides a great potential for MIR applications. Ethnic music is not only a unique environment of new timbres, rhythms and textures, it is also a fragile cultural heritage that can surely use some specialised research in order to be easier and better accessible. The aspect of innovation may be another argument for attracting new researchers, because only few tools are available and such research full of new challenges can form a new branch within the MIR community. In the end this effort seems to be necessary to achieve the goals stated in Downie's [16] definition of Music Information Retrieval: "MIR is a multidisciplinary research endeavor that strives to develop innovative content-based searching schemes, novel interfaces, and evolving networked delivery mechanisms in an effort to make the world's vast store of music accessible to all."

8. REFERENCES

[1] D. Moelants et al., Problems and Opportunities of Content-based Analysis and description of Ethnic Music, International Journal of Intangible Heritage, (2007), 58-67.

[2] G. Tzanetakis et al., Computational Ethnomusicology, Journal of Interdisciplinary Music Studies (2) 1 (2007) 1-24.

[3] S. Doraisamy et al., A Study on Feature Selection and Classification Techniques for Automatic Genre Classification of Traditional Malay Music, Proc. 9th Internat. Conf. on Music Information Retrieval, Philadelphia, USA, 14-18 September 2008, pp. 331-336.

[4] M. Wright et al., Analyzing Afro-Cuban Rhythm Using Rotation-Aware Clave Template Matching with Dynamic Programming, Proc. 9th Internat. Conf. on Music Information Retrieval, Philadelphia, USA, 14-18 September 2008, pp. 647-652.

[5] D. Moelants et al., Problems and Opportunities of Applying Data- and Audio-Mining Techniques to Ethnic Music, Proc. 7th Internat. Conf. on Music Information Retrieval, Victoria, BC Canada, 8-12 October 2006, pp. 334-336.

[6] B. Bozkurt, An Automatic Pitch Analysis Method for Turkish Maqam Music, Journal of New Music Research 37 (1) (2008) 1-13.

[7] P. Chordia & A. Rae, Raag Recognition Using Pitch-Class And Pitch-Class Dyad Distributions, Proc. 8th Internat. Conf. on Music Information Retrieval, Vienna, Austria, 23-30 September 2007, pp. 431-436.

[8] I. Antonopoulos et al., Music Retrieval by Rhythmic Similarity Applied on Greek and African Traditional Music, Proc. 8th Internat. Conf. on Music Information Retrieval, Vienna, Austria, 23-30 September 2007, pp. 297-300.

[9] K. Yoshii K. & M. Goto, Music Thumbnailer: Visualizing Musical Pieces in Thumbnail Images Based on Acoustic Features, Proc. 9th Internat. Conf. on Music Information Retrieval, Philadelphia, USA, 14-18 September 2008

[10] C.N. Silla et al., The Latin Music Database, Proc. 9th Internat. Conf. on Music Information Retrieval, Philadelphia, USA, 14-18 September 2008

[11] H. Li & M. Leman, A gesture-based Typology of Sliding-tones in Guqin Music, Journal of New Music Research, (2007), 61-82.

[12] A. Krishnaswamy, Multi-dimensional Musical Atoms in South-Indian Classical Music, Proc. of ICMPC, (2004).

[13] M. McKinney and D. Moelants, Ambiguity in Tempo Perception: What draws Listeners to different Metrical Levels? Music Perception, 24(2) (2006), 155-165.

[14] L. Naveda and M. Leman, Representation of Samba dance gestures, using a multi-modal analysis approach. Proc. of 5th Int. Conference on Enactive Interfaces, Pisa, European Enactive Network of Excellence ENACTIVE (2008) 68-74.

[15] K. Agawu, African rhythm: A Northern Ewe Perspective, 1995, Cambridge University Press.

[16] J.S. Downie, The Scientific Evaluation of Music Information Retrieval Systems: Foundations and Future. Computer Music Journal, 28(2) (2004), 12-23.

Moelants, D., Cornelis, O., & Leman, M. (2009). Exploring African tone scales. In Proceedings of the 10th International Symposium on Music Information Retrieval (ISMIR 2009), Kobe, Japan, 2009, pp. 489-494.

EXPLORING AFRICAN TONE SCALES

Dirk Moelants
Ghent University
Dirk.Moelants@UGent.be

Olmo Cornelis
University College Ghent
Olmo.Cornelis@hogent.be

Marc Leman
Ghent University
Marc.Leman@UGent.be

ABSTRACT

Key-finding is a central topic in Western music analysis and development of MIR tools. However, most approaches rely on the Western 12-tone scale, which is not universally used. African music does not follow a fixed tone scale. In order to classify and study African tone scales, we developed a system in which the pitch is first analyzed on a continuous scale. Peak analysis is then applied on these data to extract the actual scale used. This system has been applied to a selection of African music, it allows us to look for similarities using cross-correlation. Thus it provides an interesting tool for query-by-example and database management in collections of ethnic music which can not be simply classified according to keys. Next to this the data can be used for ethnomusicological research. The study of the intervals used in this collection, e.g., gives us evidence for Western influence, with recent recordings having a tendency to use more regular intervals.

1. INTRODUCTION

Scale recognition has a long tradition in the analysis of Western music. Already in medieval music theory, determining the mode and classifying pieces according to their mode was a central topic. Also in the music theories of the Middle-East and India, classification of music according to the scale (often connected to a certain 'mood') is an important topic. Not surprisingly, with the advent of computational methods, researchers started to design systems to perform the process of scale recognition automatically [1]. In recent years the focus has shifted from symbolic approaches, based on MIDI or score representations, to the analysis of musical audio files (e.g. [2 - 6]). Various systems have reached a reasonable level of success in labeling music according to the keys of the Western tonal system (cf. MIREX 2005 [7]).

Automatic analysis and classification of scales in music that is not organized according to the Western tonal system is much less developed. Some efforts that have been done to extract the scales of e.g. Australian aboriginal didjeridu music [8] or Indian classical music [9] use a reduction to Western pitch classes, thus avoiding the problems raised by irregular temperaments. Although this approach can be efficient to a certain extent, it seems limited to music with a pitch organisation that has a certain resemblance to the Western system and is problematic in terms of culture specific information. In some music the pitch-set used is as such not very important, but rather the musical gestures associated with playing or moving from one tone to another are the most characteristic aspects. This has been used in the study of Chinese guqin music [10] and Carnatic (South-Indian classical) music [11], using prototypical gestural patterns or melodic atoms to describe the melodic content of music in which the pitch is seldom stable.

Some work has been done on the scale analysis of music of the Middle-East, more precisely on Turkish [12] and Persian [13] modes. This music is characterized by the occurrence of intervals based on (roughly) a quarter tone scale. Therefore, an analysis based on a chromatic (half-tone) division of the octave can not be used. Therefore the pitch is analysed on a more continuous scale, then transformed to pitch histograms, which can be attributed to schemata that represent the modes used in this specific repertoire.

Pitch organisation in the music of Sub-Saharan Africa does not rely on a fixed theoretical framework. Ethnomusicological research has shown that a large variety of scales is in use. Often these scales use intervals that do not conform to the European chromatic scale, e.g. the use of intervals around 240 cents in (roughly) equidistant pentatonic scales [14]. However, standardized tuning systems or culture-specific classification systems do not exist. In this paper we will propose a system to explore African scales with applications in Music Information Retrieval and ethnomusicology. First we will present the collection on which the scale-detection system is applied, as well as the test-set which will be analyzed in detail. In the next chapter we give a brief description of the pitch detection and peak extraction systems used to analyze the music and how the output can be coupled with the metadata associated with the original sound files.

2. BACKGROUND

The audio set which has been used in this research, is a selection from the audio archive of the RMCA (Royal Museum for Central Africa) in Belgium. It is one of the largest collections worldwide of music from Central Africa. The audio collection consists of about 50,000 sound recordings (with a total of 3,000 hours of music), dating from the early 20th century up to now. Aiming for durable conservation and enhanced accessibility, the audio archive has been digitized entirely. Not only the audio but also accompanying metadata and contextual information have

been digitized. A database and website were developed containing complete descriptions and fragments of the audio. The results of this project can be accessed on the website http://music.africamuseum.be.

For this study a selection of 901 audio files was used. In order to be sure to get a selection that consists only of music that uses a relatively fixed tone scale (and not e.g. music for percussion ensemble), we extracted music using four common types of musical instruments: musical bow (N = 132), zither (N = 134), flute (N = 385) and lamellophone (thumb piano) (N = 250). The selection was limited to music described as solo performances, mostly they contain only the sound of the instrument, in some cases the performer also sings, accompanying him/herself on the instrument.

3. ANALYSIS

3.1 Pitch detection

The pitch algorithm used in this paper has originally been designed to perform automated transcription of sung audio into a sequence of pitch classes and their duration [15]. Original goal of this tool was the development of a query-by-humming system for retrieving pieces of music from a digitized musical library. In this original system, the acoustic signal is turned into a parametric representation of the time-frequency information. A note is assigned to the segment by identifying the highest peak in the histogram of the frame-level pitch frequencies found in the segment, and by computing the average of the pitches lying in that bin. The pitch is then converted to a MIDI note rounding the computed annotation to the closest note frequency.

For the application of pitch recognition to the study of African scales, some important adaptations had to be made. First, the time segmentation, necessary to create melodies, was not of importance for building the pitch scales and was left out. Second, the quantization of the annotations into MIDI notes was unwanted, as we want to describe music that does not necessarily follows the equally-tempered scale. Therefore, the actual frequencies were used as pitch annotations. The output of this pitch algorithm consists of a list in which every line represents 10 ms, listing six potential frequencies, each with a probability. This allows extension to polyphonic textures.

In this case however, we choose to work with largely monophonic music. Therefore only the pitch with the highest probability was retained for every 10 ms, at least if this probability was higher than a minimal threshold (in this case 0.5). Then the frequencies were transformed to a cents scale, setting C0 to zero (cents). For comparing histograms, all listed values were reduced to one octave, generating a chromavector of 1200 values, representing the scale of the piece. A typical example of the graphical representation generated by the pitch detection system is shown in Figure 1.

3.2 Peak Extraction

The pitch analysis gives us a precise representation of the pitch content used in every piece. To extract the scale, a peak analysis was performed on the histograms. As a first step, the 1200 integers are ranked by their value. Starting from the highest value, peaks are assigned. A peak is accepted if it meets all parameters. Parameters for the selection are width of the area around the peak in which another peak cannot be assigned, the size (volume) of the peak and height relative to average height. Parameters were manually optimized for this data set by trial and error. As the histograms show a large variance (small and wide peaks, high and low peaks, high and low noise levels), a mean of the best individual settings was chosen as final parameter settings (see Table 1). The analysis gives us the number of peaks, average height and a precise description of the location and size of the individual peaks (Table 2).

Parameter	Definition	Value
Places	number of lines in the input file	1200
Peakradius	width of the peaks	30
Overlap	tolerated overlap between peaks	0.25
Accept	maximal proportion: volume peak without overlap/volume peak with overlap	25
Volfact	minimal volume of a peak: volfact*(average height of histogram)*(1+2*peakradius)	1
Heightfact	minimal height of a peak: heightfact*(average height of the histogram)	1

Table 1. Parameters used in the peak detection, with the settings used in the current analysis in the Reith column.

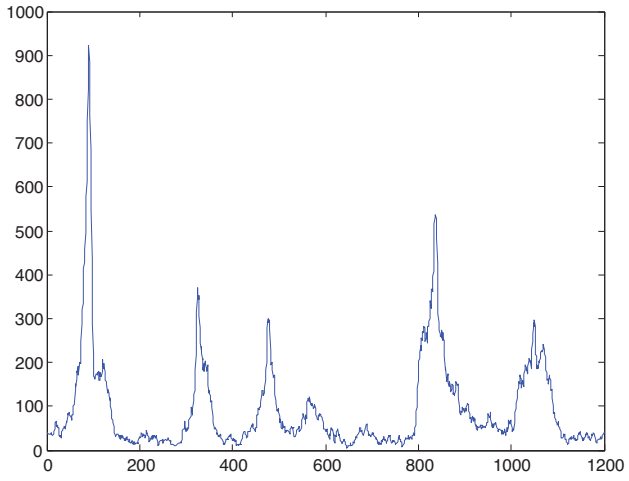


Figure 1. Example of the graphical output of the pitch analysis.

Peaks (cents)	volume	height	left side	% peak height	right side	% peak height
91	20441	922	61	0,15	121	0,21
837	17953	538	807	0,47	867	0,29
325	9603	371	295	0,13	355	0,29
476	8305	301	446	0,17	506	0,16
1050	12313	296	1020	0,57	1080	0,52

Table 2. Example of the output of the peak analysis for the piece shown in Figure 1, showing the pitches, together with information on the size of the peak.

3.3 Metadata

All the meta-data that were originally associated with the collection were digitized. Thus we get a large number of data fields from different categories: identification (number/id, original carrier, reproduction rights, collector, date of recording, duration), geographic information (country, province, region, people, language), and musical content (function, participants, instrumentation). Unfortunately, not for every recording all fields are available and often these data cannot be traced, as a large part of the collection is made up of unique historical sources.

The results of the pitch and scale analysis can be coupled with existing meta-data such as instrumentation, geographical information or date of recording. This can give us valuable information on the use of certain scales, such as geographical spread or evolution through time. As the current selection of pieces is relatively small, we used broad categories for the geographical origin (West-Africa, Southern Africa,…) and the recording time (before 1960, between 1960 and 1975, after 1975). An example of such a coupling is given in Figure 2. It gives the amount of peaks for each piece for each of the three time periods. This shows that in recent recordings, hexatonic and heptatonic scale become relatively more important while the importance of pentatonic and tetratonic scales diminishes.

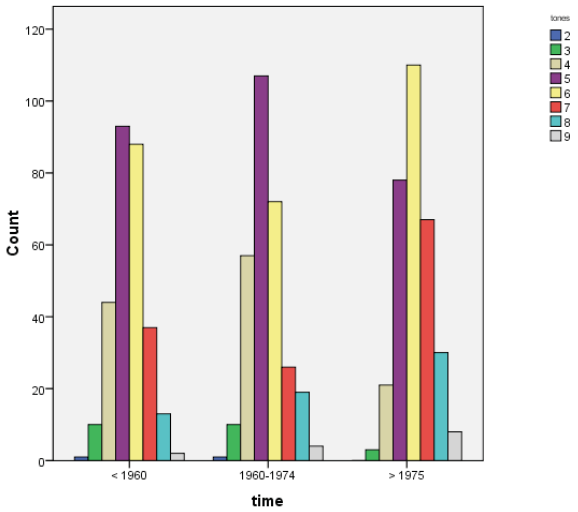


Figure 2. Bar chart representing the amount of peaks (2-9), for each the three categories of recording time: before 1960 (n = 288), between 1960 and 1975 (n = 296) and after 1975 (n = 317).

4. APPLICATIONS

The analyses made can be applied in different areas. First we will show the application of the pitch detection for data-mining applications, using cross-correlation of the pitch profiles to look for similarities. Next we will show an application of the techniques for ethnomusicological research, studying the intervals used in African scales.

4.1 Correlation analysis

The chromavectors given by the pitch analysis can be cross-correlated with each other in order to search for similar scales. As African music doesn't have a standardized tuning pitch, we need to allow a shift of pitch. Therefore, the cross-correlation technique uses every permutation of the original chromavector and returns the highest correlation from a list of 1200 correlations together with amount of cents it had to be shifted (Figure 4). Thus this method can be used for a query-by-example in which the output is a list of pieces with similar scales. This application allows to retrieve a song from a database without knowing any concrete fact about it, which is an important element for the usability of a search engine in a database of largely unknown music.

Next to this, the method can be useful for database management. The technique allows to check whether some songs are already present in their archive (so called double listed items), looking for perfect correlations without pitch shift. It can also help to establish groups of pieces with a similar origin, detecting possible links between recordings from different origins (cf. Figure 3). This could eventually lead to determination of missing meta-data.

Although the results of this analysis are promising, still some optimizations have to be done. Thus e.g. noisy pitch profiles with broad peaks, indicating less stable pitches

(e.g. from singing) (cf. Figure 3) are more likely to generate high correlations compared to pieces with very clearly defined pitches. Similarly the larger the number of peaks the more difficult it gets to obtain high correlations. Some mechanisms to deal with these differences should still be developed.

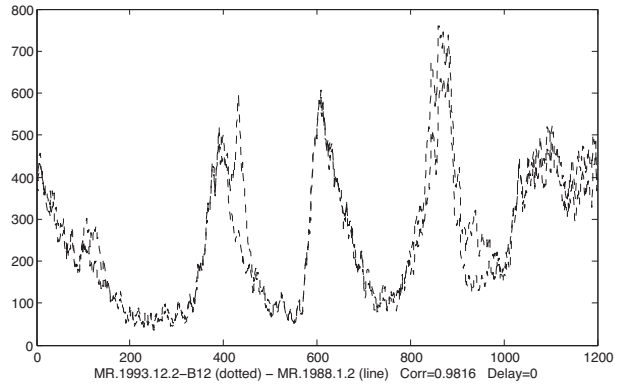


Figure 3. Graphical representation of a query-by-example, in this case the correlation is very high (r = .98) and no shift in pitch is necessary, which could point to a similar origin, despite the different sources.

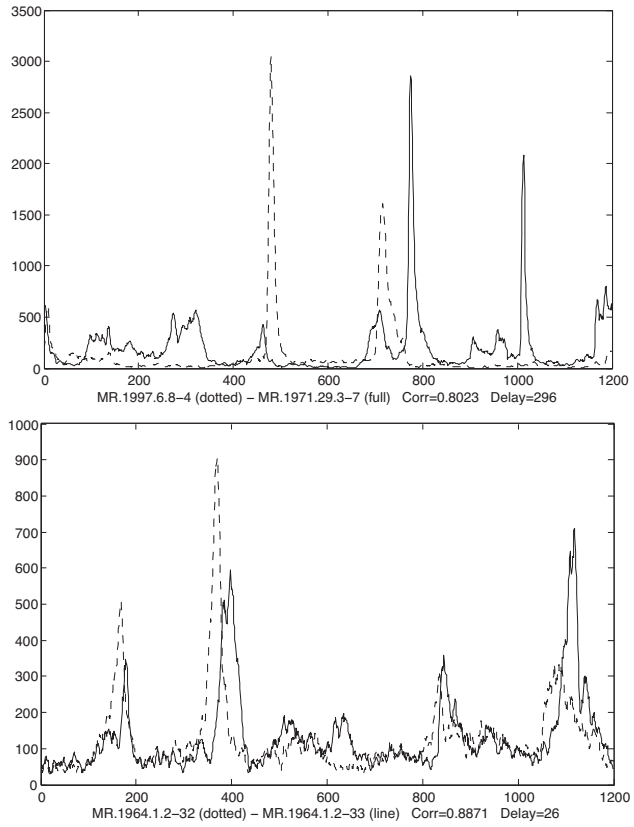


Figure 4. Two examples of a cross-correlation analysis, where the optimal correlation is found through a pitch shift. In the upper example a relatively large shift of 296 cents (about a minor third) reveals the highest similarity (r = .80), while in the lower example only a small shift of 26 is necessary to obtain the maximum (r = .89).

4.2 Interval analysis

In the analysis of 20th century Western classical music, so-called 'interval vectors' are used to express the intervallic content of a pitch-class set [16]. Using a Western chromatic scale, interval vectors are limited to an array of six numbers, expressing the amount of occurrences of each possible pitch interval (from a minor second to a tritone). With the variety of intervals found in African scales, this reduction to six numbers is not possible. Nevertheless, creating a global view on the intervals that constitute the scales can give us interesting insights in the pitch structure of the music. Are there for example any specific intervals that occur often, can we see regional differences or is there an evolution through time.

For this analysis the scales obtained from the peak analysis are transformed to an array of all possible intervals that can be built with this scale. As we work with scales reduced to one octave, the distinction between rising and falling intervals can not be made. Therefore the maximum interval size is set at 600 cents (a tritone or half an octave). For the analysis presented here, the intervals were grouped in bins of 5 cents, which gives us interval vectors of 120 elements.

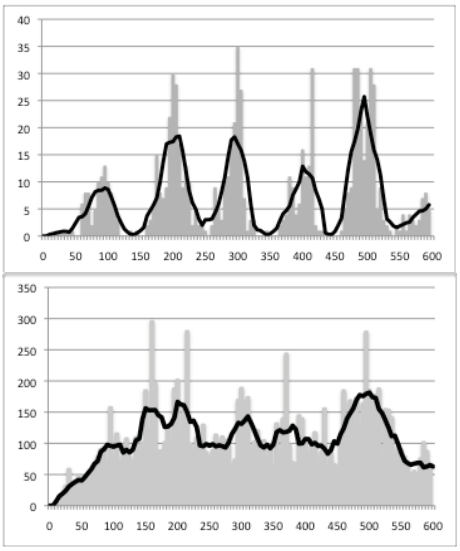


Figure 5. Comparison of all the pitch intervals found in the scale analysis from (above) the 42 pieces from the J.S. Bach's six cello suite and (below) our collection of 901 pieces of African music.

First we can make a global analysis of the intervals. In figure 5, a comparison is made between our complete collection of 901 pieces and a small sample of Western tonal music (Johann Sebastian Bach's six cello suites, played by Mstislav Rostropovich, a collection of 42 movements). In the interval analysis of the Western music we clearly see peaks corresponding to the standard intervals of 100 cents. For the African music the situation is much less clear. One similarity could be the importance of the 500 cents intervals (corresponding to the pure fourth/fifth), but the other

peaks are much less well-defined and in some cases sharp peaks appear at odd intervals (e.g. 160, 370 cents).

Now we can also couple the interval vectors with the meta-data. As an example we can look if we can find some differences in interval content between the three time periods. The analysis of the meta-data already revealed that tone scales with larger number of pitches became more important in recent recordings (cf. supra). Do we also find an influence on the pitch content? All three interval profiles are very irregular and show peaks at unexpected places, as seen in the global analysis. An interesting evolution is seen if we look at the relative share of the intervals corresponding to the Western equally-tempered scales. Counting the relative share of the 5 relevant intervals by taking the two bins around the correct interval (e.g. 95-105 cents for the minor second), we see that the share of these intervals almost doubles in the recent recordings (Table 3). Only for the minor third we don't see an increase, and the change is especially remarkable for the major seconds (also containg the minor sixths) and the pure fourths/fifths. A detailed view on the area in which pure fourths and fifths are found reveals an interesting evolution (Figure 6). The main peak seems to shift from 530 cents in the earliest recordings to 515 cents in the middle period to end up at 500 cents in the most recent recordings. This possibly also indicates a gradual evolution to a Western pure-fifth based tuning.

Interval	< 1960	1960-1975	> 1975
min. 2nd	1,46	1,87	2,20
maj. 2nd	1,57	1,71	5,20
min. 3rd	2,26	3,39	2,84
maj. 3rd	1,25	1,28	2,78
4th/5th	2,55	2,56	5,31
sum	**9,10**	**10,81**	**18,33**

Table 3. Relative share (in %) of pitches in an area of 10 cents around the Western equally-tempered intervals, for three recording time periods.

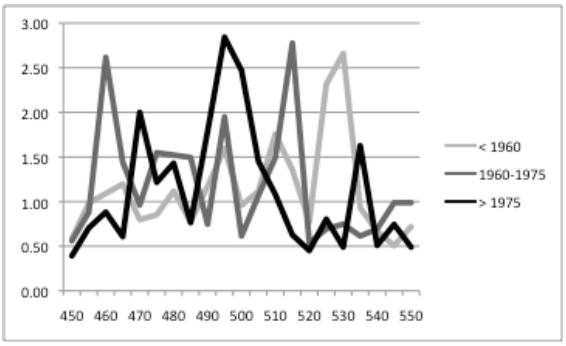


Figure 5. Relative share (in %) of pitches in bins of 5 cents between 450 and 550 cents for three recording time periods.

A detailed analysis of regional differences goes beyond the scope of this paper. Yet, we can see some interesting elements in relation to the global analysis of intervals. We see for example that the peak at 160 cents is present in every region. This shows that it is not a feature of a particular culture, but a 'pan-African' characteristic. Further ethnomusicological work is necessary to find a possible explanation for the importance of this interval. Interestingly similar interval sizes are found in the music of the Middle-East, where they are classified as 'neutral seconds' (neither minor nor major, but in between). The pitch system there however is organized according to completely different principles, so it is not clear if a direct link can be established.

5. DISCUSSION AND CONCLUSIONS

We proposed a number of methods to deal with non-standardized tone-scales, as they are found in African music. Avoiding working with a priori determined categories (such as the pitches of the chromatic scale), allows a representation and analysis of a large variety of tone scales. This was illustrated by a sample of solo-music on four different instrument types taken from the archive of the Belgian Royal Museum of Central-Africa. The results are promising, both for data-mining application and as a starting point for ethnomusicological research. Before we can expand these techniques to the whole database, several problems still have to be solved. Some important obstacles for are the presence of unaccompanied vocal music, which usually has a fluctuating pitch. This makes it very hard to extract the exact scale automatically, without applying a kind of pitch correction first. Also there is a problem with the use of percussion. The presence of percussive sounds tends to obscure the actual pitch scale used and to generate one large peak associated with the pitch of the percussion instrument. Therefore a system to suppress these percussive sounds should also be developed.

Using this relatively small sample of 901 pieces, we could already develop some methods for ethnomusicological research, creating a more elaborate view on scales and temperaments in African music in an automated way. A global comparison between the intervals found in Western and in African scales, shows that African music does not conform to the fixed chromatic scale nor has another fixed scale, however in recent recordings there seems to be a tendency to the use of more elaborate, equally-tempered scales. Further research has to be done in these historical aspects as well as on the geographical aspects of African tone scales. These techniques lead to usable applications for query-by-example, database management and classification.

6. REFERENCES

[1] H. C. Longuet-Higgins & M. J. Steedman: "On interpreting Bach," *Machine Intelligence,* vol. 6, pp. 221–241, 1971.

[2] M. Leman: *Music and Schema Theory*, Berlin, Springer Verlag, 1995.

[3] S. Pauws: "Extracting the Key from Music," in W. Verhaegh, E. Aarts & J. Korst (eds.) *Intelligent Algorithms in Ambient and Biomedical Computing*, pp. 119–132, 2006.

[4] Ö. Izmirli: "Localized Key Finding from Audio Using Nonnegative Matrix Factorization for Segmentation," in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR2007),* 2007.

[5] C. Chuan & E. Chew: "Audio key finding: considerations in system design and case studies on Chopin's 24 preludes," *EURASIP Journal on Applied Signal Processing*, Vol. 2007, No. 1, 15 pp., 2007.

[6] E. Gomez: *Tonal Description of Music Audio Signals*, Ph.D. Thesis, Universitat Pompeu Fabra, Barcelona, 2006.

[7] J. S. Downie, K. West, A. Ehmann & E. Vincent: "The 2005 Music Information Retrieval Evaluation eXchange (MIREX 2005): Preliminary Overview", in *Proceedings of the Sixth International Conference on Music Information Retrieval (ISMIR 2005),* pp. 320-323, 2005.

[8] A. Nesbit, L. Hollenberg & A. Senyard: "Towards Automatic Transcription of Australian Aboriginal Music," in *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*, pp. 326-330, 2004.

[9] P. Chordia, M. Godfrey & A. Rae: "Extending Content-Based Recommendation: the Case of Indian Classical Music," *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR2008),* pp. 571-576, 2008.

[10] H. Li & M. Leman: "A Gesture-baseed Typology of Sliding-tones in Guqin Music," *Journal of New Music Research*, Vol. 36, pp. 61-82, 2007.

[11] A. Krishnaswamy: "Multi-dimensional Musical Atoms in South-Indian Classical Music," *Proceedings of the 8th International Conference on Music Perception and Cognition (ICMPC8)*, 2004.

[12] B. Bozkurt: "An Automatic Pitch Analysis Method for Turkish Maqam Music," *Journal of New Music Research*, Vol. 37, pp. 1-13, 2008.

[13] P. Heydarian, L. Jones & A. Seago: "The Analysis and Determination of the Tuning System in Audio Musical Signals," *Paper presented at the 123rd convention of the Audio Engineering Society*, 5 pp., 2007.

[14] G. Kubik: *Theory of African Music*, Willemshaven, F. Noetzel, 1994.

[15] L.P. Clarisse, J.P. Martens, M. Lesaffre, B. De Baets, H. De Meyer & M. Leman: "An Auditory Model Based Transcriber of Singing Sequences, " *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR2002),* pp. 116-123, 2002.

[16] A. Forte: *The Structure of Atonal Music*, Yale University Press, 1973.

Cornelis O., Demey M., Leman M. (2008). 'EME: a wireless music controller for Real-Time Music Interaction', Proceedings ARTECH 2008, Porto.

EME: a Wireless Music Controller for Real-time Music Interaction

O. Cornelis[1], M. Demey[2], and M. Leman[2]

1. Faculty of Music and Drama , University College Ghent, Hoogpoort 64, Ghent, 9000, Belgium
olmo.cornelis@hogent.be

2. IPEM – Dept. of Musicology, Ghent University, Blandijnberg 2, Ghent, 9000, Belgium
michiel.demey@ugent.be, marc.leman@ugent.be

***Abstract* — In this paper an electronic sound and music controller, based on human movement using custom made wireless motion sensors, is presented. The system combines hardware technology, software applications, music theory and concepts of embodied music cognition. It enables users to sonify their gestures in real-time by creating and adapting the music they hear. The system has been demonstrated at public events, and user feedback has been collected.**

***Index Terms* — electronic music, embodiment, gesture based interfaces, music controller**

I. Introduction

Embodied music cognition [1] considers musical experience as body-mediated. The human body is thereby conceived as a natural mediator between our musical experiences (our mind) and the surrounding physical environment (e.g. musical sounds). It is claimed that the human body can be extended with technologies, so that the mind can get access to new (perhaps virtual) environments. A music instrument is a good example of such a mediation technology. Attached to the human body it allows the human mind to get into a musical reality. The task is to design music instruments in such a way that they become good mediators, that is, that they fit well with the human body, so well that the mind forgets about their mediation function and focuses on the reality which the mediator provides. The concept of an embodied mind holds the promise that it leads to new ways of analysis, creation and performance, by taking into consideration the tight connection between mind, body and physical environment.

In the present paper, the focus is on wireless motion sensors [2]. Using these sensors, the natural mediator (the body) becomes extended with a technology-based mediator that captures human movement in a very accurate way and in real-time. This allows the registration of three-dimensional manifestations of musical expression, with the promise of exploring new opportunities in music activities. Even studying social interaction and interhuman behavior becomes possible when sets of sensors are used by a group of performers and listeners.

The EME is a wireless music controller for real-time music interaction that is based on the theory of embodied music cognition. The system envisions the concrete realization of social embodied music experiences through an artistic application. The EME allows listeners/performers to act and interact directly onto sound by the expressive movements of their body. In the present version of the system, the sensors are attached onto small gloves so that they are placed on top of both hands, allowing spontaneous (re)action of the listeners/performers. Based on the natural movements and/or expressive gestures of one (or more) user(s), the hardware device allows the control of both sound (e.g. timbre) and music (structure). The result is a multi-purpose wireless music controller for real-time social music interaction that can be easily expanded towards other audio-effect generators.

II. Technical Details

The EME works with several sensors at the same time and each individual sensor steers multiple signals. The current system consists of three main parts namely the motion sensors, called HOP sensors [2], which send their data to a Max/MSP[1] patch which extracts different features from the movement and controls musical playback by Ableton Live[2] (see also [3]).

The HOP sensors are custom made, standalone, wireless, 3 dimensional accelerometers that send their data

[1] http://www.cycling74.com
[2] http://www.ableton.com

at a rate of 100 Hz using the 2,4 GHz ISM band to a dedicated receiver. This receiver is recognized in the computer as a COM port which enables the readout in Max/MSP. Each sensor has its own Li-Po battery which guaranties an operation time of 18 hours. The sensors have a dimension of 55 mm (long) x 32 mm (wide) x 15 mm (thick) which makes them very suitable for placement on the hands of the users.

The acceleration data from the motion sensors is captured in Max/MSP which extracts several parameters from this data stream in real-time. Three parameters are currently used in the EME, namely, intensity of movement, orientation and triggering. (i) The intensity of movement is the size of the jerk, which is the derivative of the acceleration signal, calculated over a sliding window of a certain number of data points. This intensity of movement quantifies the mount of changes in the acceleration of the movements and can be related to the size and force of the gesture. (ii) Orientation of the movement sensors is calculated using the constant force of the earth gravity. A tradeoff was found between stability, using low pass filters, and responsiveness, using no filters. (iii) The trigger is calculated by taking the difference of two successive acceleration samples. This is done for each of the 3 measured directions after which the absolute values are added and surpass a fixed threshold, which is defined from extensive use. The approach eliminates the offset in acceleration due to the gravity of the earth and detects sudden changes in the movement. Furthermore the fixed threshold can be set to a small value which enables the successful use of a broad spectrum of movements ranging from very small to very large movements. Trigger and orientation signals are translated to MIDI messages which control music samples played back by Ableton Live.

Ableton Live is a commercial music software, which is often used for music composition and mastering, but also for performance. By using MIDI as a control it is possible to start/stop audio loops and change various parameters of build-in effects like reverb, delay, filter, etc.

III. Sound and Music Control

A successful haptic music controller requires that performers, including non-musicians, embody their control of sound and music. Essential development issues are on the one hand the necessity that performers have to feel that they control and can steer the music with their movements and on the other hand that also non-musicians could make music with it. Therefore, different typologies of movement of performers were analyzed and mapped onto potential sound and music influencing parameters, as shown in Table 1. The current prototype of the EME is based on (i) volume triggering, (ii) timbre modulation and (iii) sample (or note) triggering.

The music composed for the EME had to be conceived

TABLE I
Summary of Distinct Hand Movements and Potential Use in EME

Movement	Calculation	Music Control	Disadvantage	Advantage
Repetitive hits (i)	Size of the jerk, the derivative of the acceleration, over a sliding window of n samples	Volume of a percussion track	Needs calibration in mapping to volume and adjusting envelope behavior in attack and sustain	Quite intuitive and good in quantization of percussive sounds
Turn around (ii)	Calculate the rotation around the 3 axis with use of gravity	Frequency of a filter	The 3 directions are coupled / sudden movements disrupt angle calculation	Intuitive (feels like turning knob) responsivity can be adjusted
Abrupt changes (iii)	Use a fixed threshold on the difference of the signal	Trigger a note or musical clip	Needs algorithmic composition to keep the melody interesting	Quite intuitive
Similar movements	Correlation of 2 signals	Adds rewarding melody	Only on or off state	Robust Nice encouragement
Periodical Movements	FFT over 4 seconds with 2 seconds overlap [3]	BPM	It takes at least 2 seconds for the tempo to adjust	Very robust to different kinds of movement

in such way that the selected types of movement (and related influence) are connected to acceptable and obvious parameters in the music. The music had to be robust and meanwhile interesting enough for both systematic and repeated movements as well as for short and irregular gestures. The musical basis for each performance is a non-stop loop that offers a basic percussion track and chords. The other layers in the music can then be triggered or modulated by movement, so that the user(s) get an instantly enriching sound. (i) Volume triggering, for example, controls the volume of another (synchronized) percussion loop according to the quantity of movement. A movement discontinuation stops the extra percussion, while a continuous movement controls the volume level according to the intensity of the gesture giving the performer the feeling of an embodied control. (ii) Timbre modulation is obtained by rotation of the hand. This can be achieved simultaneously with the volume triggering. Multiple timbre filters are possible, and even in a combination: reverb, flanger, distortion, high pass and low pass filter. (iii) Samples (or notes) can be triggered through abrupt changes in movement. The sonification is based on an algorithmic pitch component that takes into account pitch distributions related to particular musical genres, further discussed in the next section.

IV. Algorithmic Composition

Adding notes, and effectively creating a melody, to the sound tape that is playing is quite challenging when the notes are triggered based on a single trigger condition. This problem was solved using an algorithmic composition. The triggered note is chosen from a set of well defined possibilities that are calculated percentagewise by rating all the possible next candidates based on the previous note and on its time position in the actual chord. The choice of the note is in fact an organized randomness, disfavoring notes that would not sound good, and attribute a nonzero percentage to all the notes that would fit well. For each different chord in the music, a specific table of percentages has been assigned for all twelve notes within one octave, respecting the underlying chord structure of the song. These limitations in the randomness were necessary for guaranteeing the musicality, and are for each song different. An additional advantage for this individualized assignment of values is the possibility for compounding melodic lines towards musical genre. Certain melodic intervals can de banned, marginalized, favored or even obliged. Basic theoretical principles employed are rules of tonal harmony and counterpoint, but the composer is free to distribute the percentages as he wants to. It allocates the composer to assign certain rules and or limitations, but the progress of the melodic evolution cannot be foreseen. Melodic progress can thus have dodecafonic rules (percentages are evenly spread for the 12 tones), harmonic rules (Major, Minor…), modal rules, Messiaen modi, ot other personal scales…

V. Users Feedback

The EME was firstly presented at the public event Resonance, on February 16th 2008, in Ghent, Belgium. Participants could try the system with five different musical styles and were interviewed afterwards about their experiences. In general, participants were enthusiastic about the embodied way of controlling the music. Musicians and non-musicians performed just as well, both mostly with a small learning period before having grasped all the possibilities of the sensors. Particularly to have mastered the different sound effecting opportunities of one sensor simultaneously, was a demanding task which needed some practice. Nonetheless performers tried with enthusiasm and clearly improved their skills, up to a level of really controlling different aspects within the presented musical piece. Some people clearly moved strictly on the beat, resulting in a repetitive movement of the hand, while others more tend to dance on the music and thus moved in a more lyrical way, making curves and gestures with their hand. The musical output between these ‘punchers’ and ‘floaters’ differed, because different parameters were used more or less intensive.

A follow-up demonstration was given at the 14th International Conference on Auditory Display (ICAD) organized at IRCAM in June 2008, in Paris, France.

These first observations, the interviews and the analysis of video material made at Resonance, led to a schedule showing the most important elements for interference that will have development consequences.

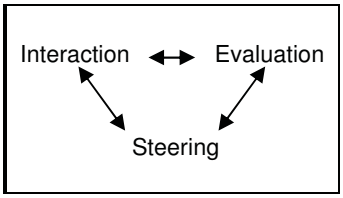


Fig. 1. Musical steering versus social interaction / Users evaluation versus musical steering / Game interaction versus scientific measurement

Depending on the goal of the researcher, the intention of the performer or the need for musicality, the emphasis of

certain aspects of the EME changes: Boosting the level of possibilities of interaction will improve game quality and level of sociability, emphasis on the musical steering will improve the outcome of composition, but will reduce the level of freedom and user-reaction quality. When the EME is used within music or social studies research, scientific measurement and capture of the data is emphasized, decreasing the importance of the musical output. This field of tension has consequences for further development, not excluding one of the parameters, but trying to take them into account in the further development of the EME such that emphasis can be steered flexible towards the needs of the context.

VI. Artistic Performance

An altered version of the EME has been used in an artistic performance, a piece for tape and voice[3][4], where an opera singer could control the tape composition through her gestures.

Fig. 2. Performer Chia-Fen Wu sings while her embodied actions influence in real-time the ongoing tape recording.

Since the tape is a fixed composition, the algorithmic composition component was not needed for this application. During the performance 2 HOP sensors were positioned at the top of the hands of the singer, attached onto small gloves. These sensors controlled a high-pass filter and the pitch of a grain delay through the orientation of the hands, effecting directly the musical tape. This gave the singer real-time control of the music to which she was singing, opening up more interaction of a tape composition with its performer, and by which every performance would generate a varying musical result. The opera singer found the experience both liberating and artistically very interesting. Bodily expressions - often gestures - ensued on the vocal part, felt natural and caused a closer blending of voice and tape.

[3] The composition Con Golese by Olmo Cornelis is a confrontation of Western singing and non-Western musical samples from the Africamuseum in Brussels, Belgium achieved by the DEKKMMA-project (http://music.africamuseum.be).

VII. Future Work

Based on a first feedback of users, we conclude that the EME has an interesting potential concept, which deserves further development in terms of music control and social interaction paradigms. We notify that some visual feedback can help people in fine controlling the timbre filters. However, adding visual elements will create more options, it will limit some other elements as for example the freedom of the performer, being shackled to a screen.

The music model will be refined towards more options and combinations of options and real-time audio effects, including timbre modification and time-stretching. This opens interesting perspectives for gesture-based transformations of the human voice.

Compositional more songs and audio will be produced in order to have an extended audio database for performers, and a composition will be written allowing all components of the EME to be used.

Acknowledgement

The authors wish to thank all users for their valuable feedback, and especially Chia-Fen Wu who performed Con Golese with the EME music controller.

This work is funded by the EmcoMetecca project Ghent University and University College Ghent.

References

[1] Leman, M. (2007). Embodied music cognition and mediation technology. Cambridge, MA: The MIT Press.

[2] Kuyken, B., Verstichel, W., Bossuyt, F., Vanfleteren, J., Demey, M., Leman, M. (2008) “The HOP sensor: Wireless Motion Sensor” (submitted)

[3] Demey, M., Leman, M., Bossuyt, F., Vanfleteren, J. (2008) “The Musical Synchrotron: using wireless motion sensors to study how social interaction affects synchronization with musical tempo” (submitted)

[4] Cornelis, O. (2007). Con Golese. Composition for Tape and Voice (unpublished)

Antonopoulos, I., Pikrakis, A., Theodoridis, S., Cornelis, O., Moelants, D., Leman, M. (2007). Music retrieval by rhythmic similarity applied on Greek and African traditional music, in: *Proceedings of Eighth International Conference on Music Information Retrieval (ISMIR2007)*, Vienna, Austria, 23–30 September 2007, pp. 297–300.

MUSIC RETRIEVAL BY RHYTHMIC SIMILARITY APPLIED ON GREEK AND AFRICAN TRADITIONAL MUSIC

Iasonas Antonopoulos, Aggelos Pikrakis and Sergios Theodoridis
Dept. of Informatics & Telecommunications
University Of Athens, Greece

Olmo Cornelis, Dirk Moelants and Marc Leman
IPEM - Dept. of Musicology
Ghent University, Belgium

ABSTRACT

This paper presents a method for retrieving music recordings by means of rhythmic similarity in the context of traditional Greek and African music. To this end, Self Similarity Analysis is applied either on the whole recording or on instances of a music thumbnail that can be extracted from the recording with an optional thumbnailing scheme. This type of analysis permits the extraction of a *rhythmic signature* per music recording. Similarity between signatures is measured with a standard Dynamic Time Warping technique. The proposed method was evaluated on corpora of Greek and African traditional music where human improvisation plays a key role and music recordings exhibit a variety of music meters, tempi and instrumentation.

1 INTRODUCTION

In the context of Music Information Retrieval (*MIR*), finding music recordings with similar rhythmic characteristics is a highly desired task both for the untrained listener and the musicologist.

Over the years, several methods have been proposed in the context of western music for retrieving music with similar rhythmic characteristics, e.g. tempo, meter and rhythmic patterns. Here a short overview of relevant papers is given: The work in [1] measures the similarity between rhythmic patterns extracted from music recordings and artificially generated percussive sounds. The approach in [2] extracts temporal patterns from the energy envelop of the signal in an attempt to classify music recordings to predefined classes. In [3] a set of classification schemes are proposed that are based on extracting rhythmic patterns from the signal's spectrum. The method proposed in [4] focuses on ballroom dances and is based on features stemming from the histogram of *Inter - Onset - Intervals*. Finally, the work in [5] evolves around self similarity analysis of the music recording. In some of the above methods, the term "rhythmic signature" is used to as a means to encode fundamental rhythmic characteristics of the music recordings.

This paper focuses on rhythmic similarity in non Western music, i.e., Traditional Greek and African music, which have so far received little attention in the field of MIR. Such traditions impose a number of research challenges, mainly due to the complexity of the music meters, the system of music intervals and the highly improvisational attitude of the music performers. These elements make it difficult to incorporate these musical genres in traditional MIR systems. The latter gives an additional research challenge as serves the preservation of cultural heritage and also highlights the importance of MIR systems to apply to corpora that fall outside the traditional western schemes. It would give the public easier access to a world of music they are not usually familiar with and it would also provide ethnomusicologists with tools that can help them to study musical genres that are mostly not available as written scores.

In an attempt to measure rhythmic similarity in such music corpora, this paper exploits the repetitive nature of the music recordings by means of Self Similarity Analysis. This type of analysis permits to reveal periodicities that are inherent in the music signal. Such periodicities are encoded in a sequence of values to which we also refer by the term *rhythmic signature*. To this end, we also investigate the possibility to apply an optional thumbnailing scheme as a preprocessing step to extracting rhythmic signatures. Similarity measurement between signatures is then performed by means of standard *Dynamic Time Warping* techniques. The main idea behind this approach is that a low matching cost corresponds to similar recording in terms of rhythmic characteristics.

Section 2 presents the proposed audio thumbnailing scheme and Section 3 describes how *rhythmic signatures* are extracted from the music signal. The proposed similarity measure is presented in Section 4. Results and implementation details are given in Section 5 and conclusions are drawn in Section 6.

2 THUMBNAILING SCHEME

The proposed audio thumbnailing scheme is optional and can be considered to be a variation of the method proposed in [6], in the sense that a different feature extraction scheme is used in this paper.

2.1 Feature extraction

At a first step, the music recording is short-time processed by means of a moving window technique. The short-term frames are chosen to be non-overlapping, $\simeq 186$ *msecs* long and are multiplied by a *Hamming* window. Each frame is given as input to a mel-scale Filter Bank ([7, 8, 9]) that consists of overlapping triangular filters, whose center frequencies, F_k, follow the equation:

$$F_k = F_0 * 2^{k/6} \quad (1)$$

Equation (1) suggests that the center frequencies of the filters coincide with the frequencies of whole tones on a chromatic scale, starting from $F_0 = 110Hz$ and moving up to $\simeq 6.3KHz$, resulting into 36 filters, which cover approximately six octaves. If $\widetilde{O_k}, \quad k = 1, \ldots, 36$, is the output of the k-th filter, then the resulting MFCCs are given by the equation:

$$c_n = \sum_{k=1}^{36} (\log \widetilde{O_k}) \cos[n(k - \frac{1}{2})\frac{\pi}{36}], \quad n = 1, \ldots, 36$$

In the sequel we will refer to this type of MFCCs as *chroma-based MFCCs* ([11]) due to the similarities it bears with the "chroma vector" [10] .

To proceed, let $\underline{c}(n)$ be the $[36 \times 1]$ vector of *chroma* based MFCCs from the n-th frame, where $n = 1, \ldots, N$, and N the number of short-term frames. Then the sequence of MFCC vectors can be written in matrix notation as:

$$\mathbf{C}_{36\times N} = [\underline{c}(1)\ \underline{c}(2)\ \ldots\ \underline{c}(N)]$$

At a next step, *Singular Value Decomposition* (*SVD*) is applied on the transpose, $\mathbf{C}^T$, of $\mathbf{C}$, i.e., :

$$\mathbf{C}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}, \quad (2)$$

where $\mathbf{U}_{N\times 36}$ holds the right-singular vectors (projection matrix), $\mathbf{\Sigma}_{36\times 36}$ is the matrix of singular values and $V_{36\times 36}$ holds the left-singular vectors of the decomposition. At a final step, the first six rows of the transpose, $\mathbf{U}^T$, of $\mathbf{U}$, are selected as the feature sequence that will be used to generate the Self Similarity Matrix (*SSM*) [5].

2.2 Thumbnail selection

As explained above, the *SSM* is generated from the first six rows of $\mathbf{U}^T$ using the Euclidean Distance function as metric. By its definition ([5]), the *SSM* is symmetric around the main diagonal and it therefore suffices to focus on it's lower triangle. At a first step, the SSM is correlated with a rectangular window, w (size $D \times D$). The window has 1's on the main diagonal and zeros elsewhere. If (i, j) are the position indices of an element of *SSM*, the upper left corner of w is chosen to coincide with (i, j). The correlation result for $SMM(i, j)$ is therefore computed as follows:

$$\begin{aligned} S(i,j) &= \sum_{d_1=0}^{D-1}\sum_{d_2=0}^{D-1} SSM(i+d_1, j+d_2)w(d_1, d_2) \\ &= \sum_{d=0}^{D-1} SSM(i+d, j+d) \end{aligned} \quad (3)$$

At a second step, let $S(k, m)$ be the lowest value of S. By the definition of the SSM, $S(k, m)$ resides on the diagonal with index $k - m$. Due to the preceding correlation step, elements $\{S(k, m), S(k+1, m+1), \ldots, S(k+D-1, m+D-1)\}$ form a segment on the diagonal that defines the desired thumbnail. The two corresponding feature subsequences, i.e. two instances of the thumbnail are $\{U_k^T, U_{k+1}^T, \ldots U_{k+D-1}^T\}$ and $\{U_m^T, U_{m+1}^T, \ldots, U_{m+D-1}^T\}$ respectively. Parameter D controls the size of the thumbnail and is user defined, depending on the corpus under study (see Section 5).

It has to be noted that the proposed thumbnailing scheme is optional, depending on the user and the needs of the experiment. If it is skipped, the rhythmic signatures (see next section) will be extracted by taking into account the whole audio recording. This is desirable if one is unsure whether the two instances of the extracted thumbnail are indeed representative of the complete music recording. In the sequel, for the sake of simplicity, the term music signal will refer to either the two instances of the selected thumbnail or the complete music recording.

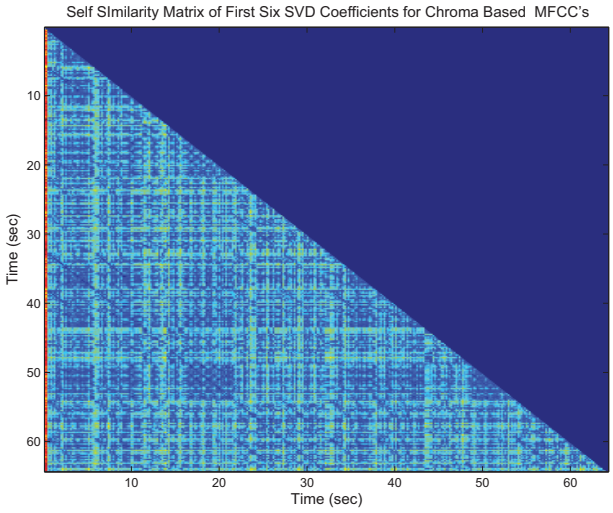


Figure 1. The *Self Similarity Matrix*, SSM, calculated for the first six coefficients of the transpose left-singular values matrix U^T stemming from the *SVD* analysis of the *chroma* based MFCCs.

3 EXTRACTING RHYTHMIC SIGNATURES

3.1 Feature extraction

At a first step, the music signal (i.e., the two thumbnail instances or the complete recording) is short-time processed

to extract a sequence of *chroma*-based MFCCs, as in Section 2.1. However, this time, shorter, overlapping windows are used (window length is $\simeq 93$ *msecs* and window step is 11.6 *msecs*). Following the notation that was introduced in Section 2, let

$$\mathbf{C} = [\underline{c}(1)\ \underline{c}(2)\ \ldots\ \underline{c}(N)]$$

be the new sequence of MFCCs. At a first step, $\mathbf{C}$ is long-term segmented with a moving long-term window (window length is 4 *secs* and step is 1 *sec*). To simplify notation, let

$$\mathbf{C_t} = [\underline{c_t}(1)\ \underline{c_t}(2)\ \ldots\ \underline{c_t}(M)]$$

be the subsequence that corresponds to the t-th long-term window, where M is the window length measured in number of frames. The *SSM* is then calculated for each long-term window, using the Euclidean Distance metric. For the t-th long-term window, the mean value, $R_t(k)$, of each diagonal in the lower SSM triangle is calculated, i.e., :

$$R_t(k) = \frac{1}{M-k}\sum_{l=k}^{M}\left\|\underline{c_t}(l), \underline{c_t}(l-k)\right\|, \qquad (4)$$

where M-k is the length of the kth diagonal, k is the diagonal index and $\|.\|$ is the Euclidean distance function.

Each R_t is treated as a signal. At a next step, the mean value, R_μ, of all R_t's is calculated, i.e., :

$$R_\mu(k) = \frac{1}{T}\sum_{t=1}^{T} R_t(k),$$

where T is the number of the long-term windows of the music signal. R_μ is then normalized to unity, i.e., :

$$R_\mu(k) = \frac{R_\mu(k)}{max(R_\mu)}$$

As can be seen in Figure 2, where R_μ is plotted against k, R_μ exhibits a number of valleys (local minima). Each valley corresponds to a periodicity that is inherent in the music signal. Such periodicities are related with the rhythmic characteristics of the recording, as the music meter and tempo [11]. Therefore, in the sequel we will refer to R_μ as the *rhythmic signature* of the music recording. The main idea behind this approach, is that, recordings with similar rhythmic characteristics are expected to yield "similar" signatures (as can be seen in the upper part of figure 2). On the contrary, different rhythmic characteristics will result into "dissimilar" signatures (bottom part of figure 2). Therefore, the next challenge is to devise a similarity measure for signatures.

4 SIMILARITY MEASURE FOR SIGNATURES

If L is the number of music recordings in a corpus, L rhythmic signatures are first extracted and stored as metadata. In order to measure similarity between signatures, a standard *Dynamic Time Warping* cost has been employed

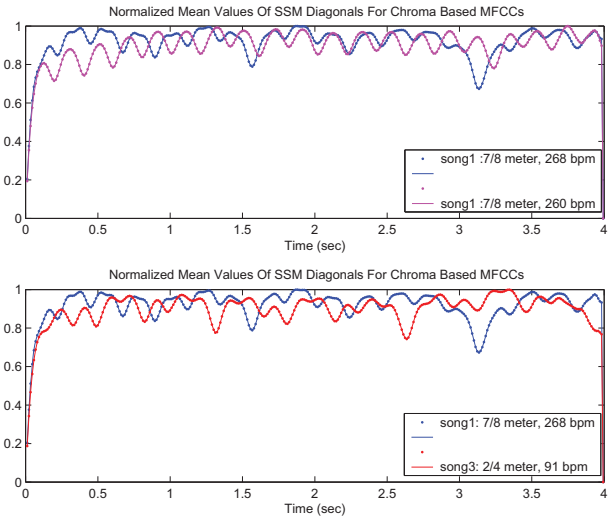


Figure 2. *Top*: *Signatures* from two recordings of music meter $\frac{7}{8}$ and *tempi* 268*bpm* and 260 *bpm* respectively. *Bottom*: *Rhythmic signature* extracted from a recording of meter $\frac{7}{8}$ and tempo 268 *bpm* and *rhythmic signature* of meter $\frac{2}{4}$ and tempo 91 *bpm*.

[12, 13]. As is the case with *DTW* techniques, a set of local path constraints needs to be first defined. In our study we experimented with two types of constraints, i.e., *Sakoe-Chiba* and *Itakura* (shown in Figure 3) and adopted the former.

If a rhythmic signature is drawn from the corpus, its matching cost against the remaining L-1 signatures is calculated using the adopted DTW technique. This procedure yields $L-1$ cost values which are sorted in ascending order, with the lowest values indicating high similarity. The next section focuses on evaluating this matching scheme on two corpora of traditional Greek and African music.

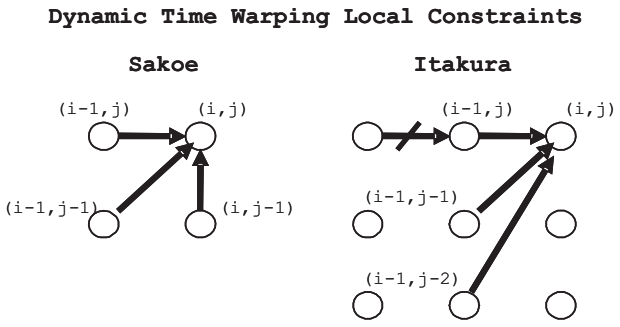


Figure 3. Local constraints used in the *DTW* algorithm for the comparison of *rhythmic signatures*.

5 EXPERIMENTS AND RESULTS

5.1 Corpus of Greek Traditional Dance music

The first corpus of our study consists of 220 tracks of Greek Traditional Dance Music, which are drawn from

various Greek regions (mainland and the islands). The tracks were manually categorised into four genres as shown in Table 1. These genres exhibit certain variety in terms of instrumentation and rhythmic characteristics, i.e., tempo varies in the range $90 - 280$ *bpm* and the following music meters are encountered: $\{\frac{2}{4}, \frac{3}{4}$, and $\frac{7}{8}\}$. It is important to notice, that by the nature of Greek Traditional music, dances of the same genre exhibit similar rhythmic characteristics, i.e., tempo and music meter. In addition, Greek dances favor the use of the proposed thumbnailing scheme because of their highly repetitive structure.

class id	# of songs	meter	tempo range (bpm)
1	53	$2/4$	91-95
2	63	$3/4$	93-105
3	62	$7/8$	250-280
4	42	$2/4$	150-180

Table 1. Description of the Greek Traditional Dance corpus.

From the corpus description it can be noticed that the longest music meter duration (approximately 2 *secs*) appears in class 2 for tempo $\cong 90$ *bpm* and music meter $\frac{3}{4}$. By using a thumbnail which is 10 *secs* long, the longest music meter is repeated up to 5 times in each long-term segment. Our study has revealed that, this expected number of repetitions is sufficient for the extraction of reliable *rhythmic* signatures and justifies our choice for the length of thumbnails. Similarly, the longest meter duration affects the long-term window length. By the definition of self similarity analysis, a periodicity of k lags will manifest itself as a valley of R_μ, if the long-term window is at least $2k$ long. Therefore the length of the long-term window was chosen equal to $4secs$ to capture the periodicity of the longest music meter. Finally, the range of lags of R_μ on which DTW is employed starts with the lag corresponding to the fastest tempo value and reaches up to the lag that corresponds to the longest meter-tempo pair.

Table 2 presents the confusion matrices for the Greek corpus in terms of class *precision* and *recall*, having applied the leave-one-out method on the whole corpus. The table reveals that limited confusion occurs between the pair classes 3 and 4 and the pair of classes 1 and 2, when only the best result (lowest matching cost) is examined. The results in Table 3 take into account the two lowest matching cost values. Table 3 reveals that the confusion between the aforementioned pairs of classes remains approximately the same within statistical confidence.

5.2 Corpus of African music

A collection of 103 pieces was selected from the music archives of the Royal Museum of Central-Africa in Tervuren (Brussels). This institute has one of the most important collections of African music in the world. This collection is currently digitized, the digital sound archive can be visited at: `http://music.africamuseum.be`. The current selection contains field recordings of Congo

Precision %	Class 1	Class 2	Class 3	Class 4
Class 1	**94.3**	3.2	1.7	0
Class 2	3.8	**96.8**	0	0
Class 3	1.9	0	**96.6**	10.9
Class 4	0	0	1.7	**89.1**
Recall %	Class 1	Class 2	Class 3	Class 4
Class 1	**94.3**	3.8	1.9	0
Class 2	3.2	**96.8**	0	0
Class 3	1.6	0	**90.3**	8.1
Class 4	0	0	2.4	**97.6**

Table 2. Precision and recall values for the four classes of the Greek Traditional Dance corpus when only the lowest matching cost is taken into account.

Precision %	Class 1	Class 2	Class 3	Class 4
Class 1	**93.5**	2.4	2.6	0
Class 2	4.7	**97.6**	0	0
Class 3	1.9	0	**95.7**	10.9
Class 4	0	0	1.7	**89.1**
Recall %	Class 1	Class 2	Class 3	Class 4
Class 1	**94.3**	2.8	2.8	0
Class 2	4	**96**	0	0
Class 3	1.6	0	**90.3**	8.1
Class 4	0	0	2.4	**97.6**

Table 3. Precision and recall values for the four classes of the Greek Traditional Dance corpus when the two lowest matching costs are taken into account.

and Rwanda recorded during the second half of the 20^{th} century [14]. Similarly to the Greek music corpus, the rhythmic structure is highly repetitive and contains a wide range of rhythmic structures, including irregular meters that are seldom found in Western music [15, 16, 17].

The corpus has been manually annotated with a (perceived) ground truth which has been used to evaluate the computer analysis. Two types of classification were made: one according to the meter, the other focusing on a selection of characteristic repetitive rhythmic patterns. The first classification, Table 4, uses four metric classes $\frac{3}{4}$, $\frac{4}{4}$, $\frac{5}{4}$ and $\frac{6}{4}$. The second Table 5, is restricted to 44 pieces which can be classified as variants of 5 prototypical patterns: short-long-long (quintuple); short-long-long-short-long-long-long (sextuple); long-short-short-long (triple1); short-long (triple2); and short-short-long (duple).

class id	# of songs	meter
1	27	$3/4$
2	26	$4/4$
3	24	$5/4$
4	26	$6/4$

Table 4. Description of the first set of the African corpus.

The thumbnail that was selected for each piece by the proposed method, often tended to contain parts of the song

class id	# of songs	pattern
quintuple	10	
sextuple	14	
triple1	8	
triple2	5	
duple	7	

Table 5. Description of the second set of the African corpus.

where the most percussive events occurred. Since this could lead to a dense regular structure that can easily be confused with patterns in which each beat is articulated the *rhythmic signatures*, R_μ's were extracted from whole audio recordings and the thumbnail scheme was skipped. Tables 6, 7 reveal that the results of both sets of African music are promising. For some groups the outcome was very good. In the future, we expect that a larger audioset will further improve the results. When manually checking the problematic cases, mistakes can mostly be related to the occurrence of variants of the main pattern within one piece. The possible variations are mostly the addition of a percussive event where a rest used to be, or the opposite: omission of a percussive event. The meter and beat stay however the same.

Precision %				
Class id	1	2	3	4
1	**68.8**	4.3	20	0
2	12.5	**82.6**	12	3.7
3	15.6	13	**64**	3.7
4	3.1	0	4	**92.6**
Recall %				
Class id	1	2	3	4
1	**78.6**	3.6	17.9	0
2	14.8	**70.4**	11.1	3.7
3	20	12	**64**	4
4	3.7	0	3.7	**92.6**

Table 6. Precision and recall values for the four classes of the first set of African corpus, when only the lowest matching cost is taken into account.

The value of the proposed algorithm at one hand is evading the time consuming labor of the manual annotations. And at the other hand the possibility for adding a valuable search parameter for musicological research and accessibility in general. This algorithm in particular uses a non-ascending, non-deterministic way of assigning measured values to musical items: A surplus value, because the perception of rhythm and tempo seems to be very personally bound and hard to objectify, especially if no scores

Precision %					
Class id	1	2	3	4	5
1	**70.80**	3.1	0	10	14.3
2	4.2	**84.4**	0	0	0
3	8.3	3.1	**86.7**	0	0
4	8.3	0	13.3	**60**	0
5	8.3	9.4	0	30	**85.7**
Recall %					
Class id	1	2	3	4	5
1	**85**	5	0	5	5
2	3.6	**96.4**	0	0	0
3	12.5	6.3	**81.3**	0	0
4	20	0	20	**60**	0
5	14.3	21.4	0	21.4	**42.9**

Table 7. Precision and recall values for the five classes of the second set of African corpus, when only the lowest matching cost is taken into account.

are available to check or rely on the "exact" value and no music theory is available for this kind of music.

6 CONCLUSIONS

This paper presented a music retrieval method based on rhythmic similarity measurement. The method yielded satisfactory results on coprora of traditional Greek and African music. In our future work, more sophisticated DTW techniques will be used on larger corpora and the possibility to extract multiple rhythmic signatures per music recording will be investigated.

7 REFERENCES

[1] J. Paulus and A. Klapuri, "Measuring the similarity of rhythmic patterns.", in Proceedings of ISMIR, Paris, France, September 2002.

[2] S. Dixon, F. Gouyon, and G. Widmer, "Towards characterisation of music via rhythmic patterns.", in Proceedings of ISMIR, Barcelona, Spain, 2004.

[3] Geoffroy Peeters, "Rhythm Classification Using Spectral Rhythm Patterns", in ISMIR 2005 London, September, 2005.

[4] F. Gouyon and S. Dixon, "Dance music classification: a tempo-based approach.", in Proceedings of ISMIR, Barcelona, Spain, 2004.

[5] Jonathan Foote, Matt Cooper, and Unjung Nam, "Audio Retrieval by Rhythmic Similarity", in Proceedings of ISMIR, Paris, France, September 2002.

[6] Bartsch, M. and Wakefield, G. H. "Audio thumbnailing of popular music using chroma-

based representations", IEEE Transactions on Multimedia, 7(1), 96-104, 2005.

[7] Malcolm Slaney, "Auditory Toolbox Version 2 #1998-010", in *Interval Research Corporation*, 1998

[8] Lawrence Rabiner and Bing-Hwang Juang *Fundamentals Of Speech Recognition*. Prentice Hall, US Edition, 1987.

[9] John R. Deller Jr, John G.Proakis and John H.L.Hansen, *Discrete-Time Processing Of Speech Signals*, Prentice Hall, US Edition, 1993.

[10] G. H. Wakefield, "Mathematical representation of joint time-chroma distributions", in *SPIE*, Denver, Colorado, 1999.

[11] Aggelos Pikrakis, Iasonas Antonopoulos and Sergios Theodoridis, "Music Meter and Tempo Tracking from raw polyphonic audio", in Proceedings of ISMIR, Barcelona, Spain, 2004.

[12] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 26, February 1978.

[13] S. Theodoridis and K. Koutroumbas, *Pattern recognition*, Academic Press, 3d Edition, 2006.

[14] O. Cornelis et al., "Digitisation of the ethnomusicological Sound Archive of the Royal Museum for Central Africa.", in IASA Journal, pp 35-43, 2005.

[15] O. Cornelis et al., "Problems and Opportunities of Applying Data- & Audio-Mining Techniques to Ethnic Music", Journal of Intangible Heritage, (in press), 2007.

[16] H. Myers,R. Widdes, *Historical Ethnomusicology (chapter in Ethnomusicology)*, Norton and Company, London, 1992, ISBN 0-393-033775-5.

[17] J.M. Chernoff , *African Rhythm and African Sensibility*, Chicago Press, 1979, ISBN 0-226-10344-7.

Cornelis, O. (2011). Theoretisch en artistiek onderzoek naar de symbiose van westerse en niet-westerse muzikale idiomen, situering. *ARIP: Artistic Research In Progress* (pp. 90–99). Gent: School of Arts.

Theoretisch en artistiek onderzoek naar de symbiose van westerse en niet-westerse muzikale idiomen, situering.

Olmo Cornelis, olmo.cornelis@hogent.be

2011

Abstract

Dit onderzoek tracht een stap te vormen in het ontwikkelen van een methode om etnische muziek te beschrijven op computationele wijze waarbij de klank als fysisch signaal geanalyseerd wordt. Deze methode kan bestaande etnomusicologische onderzoeksmethodes verrijken maar biedt ook artistieke opportuniteiten. De aangeleverde muziek-inhoudelijke beschrijvingen kunnen immers enerzijds een nieuwe kijk op de eigenheid van etnische muziek opleveren en anderzijds een inspiratiebron vormen voor compositorisch idioom.

1 Inleiding

Etnische muziek vormt een belangrijk aandeel binnen het culturele erfgoed. Sedert meer dan 100 jaar worden via musicologisch veldwerk geluidsopnames gemaakt die doorgaans bewaard worden op de originele geluidsdragers in musea. Sinds enkele jaren wordt steeds meer aandacht besteed aan duurzame bewaring en verbeterde toegankelijkheid van deze archieven door de collecties te digitaliseren. Om tot een werkelijk verbeterde toegankelijkheid te komen, is het echter nodig om doorzoekbare metadata en muzikale beschrijvingen aan te bieden. Een belangrijke stap is signaalanalyse waarbij via Music Information Retrieval (MIR) technieken het audiosignaal erg precies kan geanalyseerd worden. Het biedt bovendien de mogelijkheid een kader te creëren om etnische muzikale beschrijvingen minder te verbinden met begrippen en concepten uit de Westerse muziektheorie.

Echter blijkt dat het onderzoeken van digitaal auditief materiaal niet zo evident is. Klankanalyse kan via signaalanalyse erg precies worden uitgevoerd, maar de data die eruit voortkomen zijn niet altijd even makkelijk te interpreteren. De grote voordelen tegenover het menselijk gehoor, wat transcriptie en annotatie betreft, zijn tweeledig: precisie en automatisering. Voor grote collecties onontbeerlijke troeven. De gebruikte techniek valt onder MIR, een methode om geautomatiseerd kenmerken uit muziek te halen. Met de gigantische hoeveelheden muziek die tegenwoordig online terug te vinden zijn, kan dergelijke zoekopdracht de luisteraar op paden brengen die hij niet kende, of op het juiste moment het juiste lied te bezorgen.

Doorheen de geschiedenis werden kunstenaars vaak aangetrokken tot niet-westerse culturele elementen. Eclectisch zoekend verrijkten componisten zoals Mozart, Dvořák, Kodály, Bartók, Stravinsky, Ligeti, Reich... het muzikaal palet van hun toontaal met ritmes, melodieën en timbres.

De huidige, door digitalisering, verhoogde toegankelijkheid tot auditief materiaal, speelt in de kaart van de mogelijkheden van de hedendaagse componist, uitmondend in multi- en interculturele trends: elementen uit andere culturen worden geïmplementeerd, geïntegreerd, gecombineerd of zelfs gesynthetiseerd tot een nieuwe toontaal. De muzikale parameters en technieken die daarvoor in aanmerking komen zijn uitgebreid: toonsystemen, ritmische systemen, timbre, citaten, semantische elementen, instrumentatie... Daarenboven zorgen de digitale mogelijkheden van geluidsbewerking en montage voor een waaier aan opportuniteiten om muzikale toontalen te ontginnen.

Het eigen artistieke luik van het onderzoek zal via uiteenlopende composities trachten de verschillende mogelijkheden tot stijlvermenging af te toetsen; soms expliciet en confronterend, dan weer onopvallend en versmolten, om te leiden tot een (eigen)/ (zinnige) toontaal als componist.

2 Probleemstelling

MIR biedt een innovatieve mogelijkheid om klank te analyseren, de data die het oplevert dient grondig geanalyseerd en bestudeerd te worden vooraleer daaruit een zinvolle interpretatie kan voortvloeien. Dit vormt op zich reeds een grote moeilijkheid; onvolledig of fout interpreteren van data leidt tot misleidende of foute beschrijvingen. Als de veelbelovende aanpak van MIR losgelaten wordt op niet-westerse muziek, dan wordt op eenzelfde manier data gegenereerd, maar de grote moeilijkheid is dan interpretatie. Traditionele etnomusicologie beschreef zijn onderzoeksobject altijd met een heel erg westerse bril: het muziektheoretisch kader dat in Europa gedurende eeuwen evolueerde en vorm kreeg, werd eveneens gehanteerd om niet-westerse muzikale expressie te beschrijven. Begrippen en theorieën die de eigenheid van etnische muziek niet konden vatten, maar waarvoor geen alternatief werd gevonden om preciezer te werk te gaan. De veelal orale culturen en hun diversiteit vielen niet te vangen met éénzelfde westerse muziektheorie. Ondanks vele pogingen om traditionele muziek toch precies genoeg te beschrijven werd zelden afgeweken van het Europese muzikale denken. Bvb bleven de gevestigde twaalf toonhoogteklasses de referentie, met hooguit een melding van afwijking, maar nooit met de idee om het concept van de twaalf toonhoogteklasses op de helling te zetten. Op metrisch vlak werd uitgegaan van het westerse divisieve denken waarbij een compositie een vorm bevat, met melodie, die bestaat uit maten en verdeelbaar is in tellen. De praktijk toont echter dat etnische muziek vaak additief georganiseerd is; niet vanuit de grootste eenheid onderverdelen, maar vanuit de kleinste cel, de tel, een ostinato op te bouwen. Over timbre kan men opmerken dat geheel anders wordt omgegaan met kleuren en betekenissen. Enerzijds is het Westers begrippenkader ivm timbre behoorlijk beperkt waarbij snel wordt overgeschakeld naar beschrijvende adjectieven. Etnische muziek heeft doorgaans een minder complexe harmonische structuur waardoor minieme timbre verschillen veel belangrijker worden. Het anders omgaan met harmonie laat tevens de mogelijkheid

om een erg diverse intervalorganisatie te hanteren.

Gelijktijdig met het wetenschappelijke onderzoeksgedeelte, evolueert het artistieke luik dat onderzoekt wat de mogelijkheden zijn om tot symbiose te komen. Enerzijds stimuleert de voortdurende onderdompeling in Afrikaanse muziek de creativiteit en reikt het timbres, ritmes en vormen aan die waardevol kunnen zijn in het zoeken naar een innovatief klankidioom, anderzijds is het compositieproces een alternatieve manier om met muzikale parameters om te gaan waardoor je als onderzoeker ook tot een andere manier van kijken en luisteren kan komen.

3 Historische plaatsing

Afrikaanse muziek werd jarenlang verguisd door Europese onderzoekers. Een belangrijke reden hiertoe is dat muziek steevast samenging met dans en ritueel, hetgeen niet te verenigen was voor een missionerend ingesteld Europa dat het Christendom wenste te verspreiden. Een extreem voorbeeld noteren we bij Peter Laurent de Lucques die over Afrikaanse dans in 1705 schreef: “Er kunnen zeker zaken neergeschreven worden over dans, dans die steeds weer obsceen is. Maar ik wil daar niet over praten. De pen van een religieus man weigert dergelijke zaken op papier te zetten.” Livingstone liet tijdens zijn vele expedities in Congo zelfs enkele instrumenten vernietigen en Schweinfurth beschreef Afrikaanse muziek als het lawaai van dolgeworden katten. Stilaan kwam er ommekeer, allereerst in de plastische kunsten, waar bvb de Afrikaanse gestileerde maskers letterlijk werden overgenomen bij kubistische kunstenaars zoals bvb Picasso die op overtuigende manier ‘Les demoiselles d’ Avignon’ schilderde. Muziek werd vanaf eind 19^{e} eeuw werkelijk onderzocht, eerst onder de noemer comparatieve musicologie, waarbij vooral het cultuurvergelijkende aspect essentieel was en na WOII, onder toedoen van Jaap Kunst, onder de neutralere noemer etnomusicologie waarbij vooral de "social and cultural aspects of music and dance in local and global contexts" werden bestudeerd. In de jaren 1950-1970 werd aanzienlijk wat veldwerk verricht doorheen het Afrikaanse continent. Het opgenomen materiaal werd onderzocht en werd gehanteerd om een systematiek uit te werken bij Alan Lomax bvb voor de menselijke zang. Bruno Nettl claimt terecht dat etnomusicologie steeds weer een westers kijken naar en beschrijven van niet-westerse muzikale expressie is. In 2007 lanceert George Tzanetakis de term Computational Ethnomusicologie en zet daarmee een nieuwe trend in onderzoeksmethodologie om niet-westerse muziek te bestuderen waarbij signaalanalyse de westerse kijk moet minimaliseren.

4 Methodologie

Om niet-westerse muziek te analyseren en om afstand te kunnen nemen van het westers muziektheoretisch kader, moet er enerzijds geopteerd worden voor een zo objectief mogelijke meting en anderzijds gewaakt worden over de interpretatie die we slechts vanuit ons eigen kennis kunnen uitvoeren. Deze waakzaamheid om niet te vervallen in een westerse denkpiste is niet evident. Het blijft een steeds in vraag stellen van de genomen beslissingen richting interpretatie. De uitkomst hoeft niet dadelijk een nieuwe of alternatieve muziektheorie te zijn voor etnische muziek, maar kan hopelijk

een aanzet vormen om deze muziek correcter te omschrijven en begrijpen. De keuze van analysemethode is op zich geen groot vraagstuk: MIR biedt informatie over het fysieke signaal; welke frequenties aanwezig zijn en hoe ze evolueren in de tijd. Grootste problematiek is het optimaliseren van deze algoritmes, want het kluwen dat klank heet, is een uiterst boeiend maar complex gegeven. Zeker bij polyfone muziek wordt de versmelting van alle tonen en timbres overheen de tijd zodanig complex dat ook hierin een bijzondere uitdaging te vinden is.

Praktisch werd geopteerd voor een multidisciplinaire aanpak waarbij enerzijds een musicoloog/componist en anderzijds een computeringenieur hun aandeel vervullen in het onderzoek. Samen legden zij de basis van TARSOS, een platform voor audio-analyse dat een minutieuze precisie in klankanalyse nastreeft. De software is vrij beschikbaar op http://tarsos.0110.be en wordt uitvoerig besproken in het artikel over het onderzoek van Joren Six.

5 De collectie

Het Koninklijk Museum voor Midden-Afrika in Tervuren (KMMA) heeft een belangrijke audiocollectie opgebouwd doorheen de twintigste eeuw. Het verzamelen van geluidsdocumenten startte in 1911 met voornamelijk opnames uit Centraal-Afrika, maar werd sindsdien uitgebreid met traditionele en volksmuziek uit de hele wereld op erg diverse – en soms intussen historische – geluidsdragers zoals wasrollen, Sonofil-draadopnamen, magneetband, vinylplaat, audiocassettes, DAT of cd. De collectie omvat ongeveer 5.000 uren aan muziekopnamen, waarvan het gedeelte Afrika ongeveer 2.550 uren omvat. DEKKMMA (Digitalisatie van het Etnomusicologisch Klankarchief van het Koninklijk Museum voor Midden-Afrika) was het eerste grote digitaliseringsproject van het KMMA en beoogde de digitalisering van het hele klankarchief. De audio werd gedigitaliseerd volgens standaarden voorgeschreven door de International Association of Sound and Audiovisual Archives (IASA) en wordt bewaard op een RAID5E systeem waarbij de harde schijven beschermd zijn en gegevens opnieuw kunnen worden samengesteld in geval een schijf defect raakt.

Naast de digitalisering van de klank, dienden ook de metadata (de beschrijving van de audio) gedigitaliseerd te worden. Daartoe werd een databank en invoerprogramma ontworpen. Voor het ontwerp werd vertrokken van de structuur en de semantiek van de aanwezige traditionele papieren steekkaarten. Het grote voordeel van een 'databank op maat' is dat het beter afgestemd is op specifieke noden: snellere invoer van en toegang tot de gegevens, controle op onderhoud en consistentie van de databank, mogelijkheid tot meertaligheid, de creatie van specifieke etnische beschrijvingsvelden en invoer van aanpassingen voor ondersteuning van musicologisch onderzoek. Het ontwikkelen van een website verhoogde de toegankelijkheid tot het archief (http://music.africamuseum.be). Niet alleen (fragmenten) audio en (alle) metadata zijn langs die weg beschikbaar, maar ook de nodige contextuele informatie om het archiefmateriaal te voorzien van een cultureel kader.

Fig. 1: Instrumenten uit het Afrikamuseum: harp, hoorn, bel en trommel.

Digitalisering leidt tot een verhoogde toegankelijkheid tot muziek voor de gebruiker, die snel en eenvoudig zijn keuze moet kunnen terugvinden. Daartoe kan flexibele bevraging een oplossing bieden. Een flexibel zoeksysteem impliceert dat niet enkel naar resultaten gezocht wordt die exact beantwoorden aan de zoekcriteria, maar ook naar resultaten die bij benadering beantwoorden aan de vraagstelling, wat in het bijzonder nuttig is indien er geen enkel resultaat exact voldoet aan alle criteria. Bovendien moeten niet alle criteria scherp gedefinieerd zijn, maar is er bij het opgeven van de zoekcriteria enige flexibiliteit toegestaan (bijvoorbeeld dichter aanleunend bij de natuurlijke taal).Voor het KMMA werd een methode ontwikkeld waarmee je het archief kan doorzoeken op basis van instrumentarium. Hierbij wordt er in het klankarchief gezocht naar opnames met een gelijkaardig klinkend instrumentarium, rekening houdend met het feit dat sommige instrumenten sterk op elkaar gelijken. Bij het onderhoud van een digitaal klankarchief is enige omzichtigheid geboden. Summier en in een niet-exhaustieve opsomming, dient nauwlettend gelet op 1) de duurzaamheid van de digitale drager (zogenaamde 'digitale broosheid'), 2) de connectie tussen digitale metadata en audio, 3) de bewaring van de beschrijving van de keuzes van het (doorgevoerde of geplande) digitaliseringsproces (zogenaamde 'preservation of the preservation'), 4) de leesbaarheid van gekozen digitale formaten, 5) de uitwisselbaarheid van de gegevens van de databank (belang van exportmogelijkheden van gegevens via een gestandaardiseerd formaat, daar steeds meer databases van verschillende musea gekoppeld zullen worden tot grote online databanken). Hiermee is ook de uitwisselbaarheid met andere archieven op langere termijn verzekerd en wordt de ontsluiting van het eigen archief bevorderd.

6 Music Information Retrieval

Dankzij het Music Information Retrieval-onderzoek wordt het mogelijk om naar muziek te zoeken op basis van muzikale inhouden. Met MIR wordt immers bedoeld: geavanceerde technologie om, via analyse van klank en partituur, muzikale kenmerken te ontginnen, zodanig dat specifieke muziekbeschrijvingen, inhoudsgebaseerde muziekzoeksystemen en vernieuwend musicologisch onderzoek ontwikkeld en gestimuleerd worden. MIR levert een methodische en alternatieve kijk op audio op en kan grosso modo op drie niveaus plaatsvinden:

1) Analyse van het fysieke klanksignaal, mogelijk door analyse van digitale audio. Immers, muziek in analoge vorm met zijn continue signaal en daardoor aan subjectievere normen onderworpen beschrijving, wordt na digitalisering discreet en daardoor meetbaar op microniveau, hetgeen (beperkte) menselijke auditieve analyses sterk verrijkt, zowel kwantitatief als kwalitatief.

2) Analyse op symbolisch/semantisch niveau, waarbij de digitale partituur statistisch geanalyseerd wordt.

3) Analyse op contextueel niveau, bijvoorbeeld door web-tagging, een methode waarbij muziekbeschrijvingen aangeleverd worden door het kwantitatief verzamelen van webdata.

Het merendeel van de MIR toepassingen richt zich op commerciële muziekgenres, met als gevolg dat de bestaande toepassingen zelden geschikte beschrijvingen aanleveren voor niet-commerciële muziek. Enerzijds is de software immers geoptimaliseerd voor specifieke genres, anderzijds vraagt etnische muziek om geheel andersoortige beschrijvingen van klankkenmerken die niet zomaar te vatten zijn in het klassieke muziekjargon. Het gevaar van een dergelijke gesloten keten bestaat er in dat mensen die zoeken naar muziek altijd terechtkomen bij Westerse commerciële muziek en dat niet-commerciële muziek minder snel teruggevonden wordt. MIR-onderzoek kan geleverd worden op alle muzikale parameters zoals toonhoogte, toonladders, timbre, tempo en ritmische patronen. Voor het DEKKMMA project werden voor elk van de duizenden geluidsopnames van het KMMA archief muzikale gegevens automatisch geëxtraheerd en vervolgens verwerkt tot concrete datasets en een grafische weergave. De unieke precisie van de muzikale data toont kenmerken van etnische muziek die vroeger erg moeilijk te meten waren. De databank kan nu doorzocht worden op een manier die voorheen niet mogelijk was, namelijk op basis van muzikale inhouden. Het broze etnisch-culturele erfgoed kan op die manier meer bekendheid en toegankelijkheid krijgen. MIR opent bovendien enorme mogelijkheden voor innovatief musicologisch onderzoek.

7 Artistieke toepassing

Gedurende de eerste jaren van het onderzoek werden enkele projecten gelanceerd. **Con Golese** voor tape en zang vormt een confrontatie tussen gesamplede fragmenten van het KMMA archief en westerse zang. De verhouding van de twee partijen lijkt soms een versmelting waarbij de luisteraar nauwelijks verschillen waarneemt tussen tape en uitvoerder, maar zoekt ook andere uitersten op waarbij de confrontatie een ware oppositie en strijd is van stemmen en zangstijlen. De meeste samples werden digitaal bewerkt en vervormd waardoor een gevoel van vervreemding ontstaat, van onaards gegrol tot een idyllisch zwevend klanktapijt. In een tweede versie van Con Golese werd een technisch aspect toegevoegd waarbij de zangeres draadloze sensoren aan de hand heeft, waarmee ze de effecten en het volume van de tape kan aansturen. Het zorgt voor een intiemere en flexibelere setting daar de tape doorheen elke uitvoering toch anders klinkt op basis van de bewegingen van de zangeres.

Na Con Golese werd de gebruikte technologie ook afzonderlijk verwerkt in EME, Embodied Music waarbij gebruikers met de sensoren aan de polsen extra muzikale lijnen (percussie, melodie, bas) konden toevoegen aan de basistape.

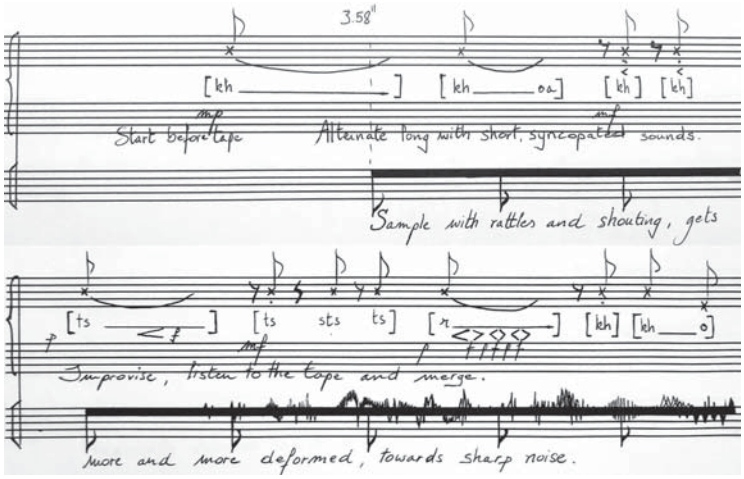


Fig. 2: Partituurfragment uit Con Golese

In **Nelumbo**, Lelie, werden 4 scènes ontwikkeld waarbij de muziek de projectie is van een lelie doorheen de seizoenen. Winter, met een ijzig kalm wateroppervlak waar hooguit een rimpeling de waterlijn breekt; lente, een knop die oprijst uit de diepe wateren en gespannen wacht tot; zomer, deze uiteindelijk krachtig openbreekt. Na het verdwijnen van de bloem, zakt alles stilaan weer weg onder het wateroppervlak, herfst. De setting bestond uit 2 sopranen en een danseres voorzien van sensoren. Al dansend kon zij in realtime de zang aanpassen door een terts- en of kwinttoevoeging. Het tweestemmig stuk kon dus uitgroeien tot een zesstemmig stuk in parallelle stemmen. De parallelle verdubbelingen zijn een duidelijke verwijzing naar de typische Afrikaanse harmonisatie waarbij geen contrapuntische lijnen worden uitgewerkt, maar waarbij terts-, kwart- en kwintverdubbelingen gebruikt worden. Dergelijk parallel verdubbelen is een gevolg van het gebruik van toontalen waarbij een ander interval voor een andere betekenis zorgt. Contrapuntische melos zou dus de betekenis in de verschillende stemmen veranderen.

Fig. 3: Scène uit Nelumbo

Traces from Tracey hanteert een mobiele compositievorm waarbij de bezetting, die willekeurig is, doorheen fragmenten partituur laveert op zoek naar samenspel. De bedoeling is dat iedereen die de ruimte binnenkomt, kan/moet participeren in de uitvoering. Er is geen publiek, er is geen traditionele 'westerse' spanningsverhouding tussen podium en zaal. Iedereen vormt een deel van de performers. Het idee berust op het muzikale gebeuren in Afrikaanse context waarbij er geen publiek is, maar waarbij iedereen deel uitmaakt van de uitvoering. Muziek vanuit een sociale cohesie. De muzikale ideeën op de partituur zijn verwerkingen afgeleid van typische Afrikaanse muzikale groepsessies. Veel percussie in een ostinate herhaling die uiteengerafeld is in verschillende partijen, met daarboven erupties van melodische of ritmische aard. Het klankspektakel is een soort performance waarbij bezoekers in en uit de zaal kunnen gaan, iedereen neemt deel voor de periode die hij wenst deel te nemen. Iedereen zal ook door zijn /haar eigen achtergrond een andere inbreng hebben in het geheel. Er is geen individualisme in de rol van solist, maar iedereen heeft wel een eigen individuele rol door de mobiele partituurvorm waarbinnen het stuk evolueert.

8 Bevindingen

Vermenging van stijlen en culturen brengt ook vragen met zich mee; kunnen/mogen stijlen en culturen vermengd worden zonder hun eigenheid en authenticiteit te verliezen, zijn er (moeten er?/mogen er?) bepaalde regels vastgelegd worden, wat zijn voor de componist de nieuwe muzikale mogelijkheden door de digitale wereld? Leidt interculturele vermenging tot vervaging van decennia lang gegroeide culturen of brengt het net redding voor kwetsbare orale tradities? Leidt vermenging tot verminking? Al deze onderzoeksvragen staan centraal in een onderzoek dat de muzikale mogelijkheden en onmogelijkheden van interculturatie door huidige digitale media bestudeert.

Het onderzoek dat ik wil voeren is enerzijds een theoretisch-filosofische benadering en anderzijds een praktisch-compositorische realisatie van de mogelijkheden en onmogelijkheden van integratie van verschillende muziekculturen. Hoe kan muziek van verschillende culturen samengebracht worden zonder één van beide conceptuele denkkaders oneer aan te doen, en tegelijk toch hoge artistieke en hedendaagse esthetische normen te hanteren die het uiteindelijke kunstwerk een esthetische en inhoudelijke meerwaarde geven.

BIBLIOGRAFIE

- Antonopoulos I., Cornelis O., Moelants D., Leman M., Pikrakis A. (2007). '*Music Retrieval by Rhythmic Similarity applied on Greek and African Traditional Music'*, Proceedings, ISMIR.

- Casey M.A., et al. (2008). Content-based MIR: current directions and future challenges, Proceedings of IEEE 96 (4) 668–695.

- Cornelis O., Lesaffre M., Moelants D., Leman M. (2010). '*Access to ethnic music: advances and perspectives in content-based music information retrieval'*, Signal Processing, Elsevier.

- Cornelis O. (2010). *'Het etnomusicologisch klankarchief van het KMMA, digitaliseren en beyond...'*, in Handboek Muzikaal Erfgoed (Ed. Schreurs E.), Resonant.

- Cornelis O., Moelants D., Leman M. (2009). *'Global access to ethnic Music: the next big challenge?'* Proceedings ISMIR, Kobe, Japan. http://www.columbia.edu/~tb2332/fmir/Papers/Moelants-fmir.pdf

- Cornelis O. (2005). *'Digitisation of the Ethnomusicological Sound Archive of the RMCA'*, IASA Journal.

- De Hen F.J. (1967). 'De muziek uit Afrika', Bulletin d'information de la coopération au développement.

- De Mey M., Cornelis O., Leman M. (2008). *'The IPEM_EME: a Wireless Music Controller for Real-time Music Interaction'*, Proceedings ICAD, Paris.

- Moelants D., Cornelis O., Leman M. (2009). *'Exploring African tone scales'*, Proceedings ISMIR, Kobe, Japan.

- Moelants D., Cornelis O., Leman M., Matthé T., Detré G., Hallez A., De Caluwe R., & Gansemans J. (2007). *'Problems and Opportunities of Applying Data- & Audio-Mining Techniques to Ethnic Music'*, Journal of Intangible Heritage.

- Nettl B. (2005). The Study of Ethnomusicology, 31 Issues and Concepts, University of Illinois Press, ISBN 0252072782.

- Schneider A. (2006) *Comparative and systematic musicology in relation to ethnomusicology: a historical and methodological survey*, Ethnomusicology 236–258.

- Tzanetakis G. (2007). et al., Computational ethnomusicology, Journal of Interdisciplinary Music Studies 1 (2) 1–24.

Cornelis, O. (2010). Een digitaal klankarchief, en nu? Over de opportuniteiten van een digitale omgeving aan de hand van het digitaliseringsproject. *Achter de muziek aan: muzikaal erfgoed in Vlaanderen en Nederland* (pp. 280–284). Acco, Uitgeverij.

Een digitaal klankarchief, en nu? Over de opportuniteiten van een digitale omgeving aan de hand van het digitaliseringsproject DEKKMMA

Olmo Cornelis

Vele klankarchieven zetten de laatste jaren hun eerste stappen richting digitalisering. Deze uitdaging dient een dubbel doel: enerzijds een goede conservatie van het originele klankmateriaal op een duurzame geluidsdrager en anderzijds het klankarchief toegankelijker maken, zowel voor een wetenschappelijk als voor een groot publiek. Hierdoor wordt de valorisatie van deze museumcollecties een feit. Door dit eenmalige digitaliseringsproces worden heel wat nieuwe onderzoeksmogelijkheden gestimuleerd.

De digitalisering van het klankarchief van het KMMA

Het Koninklijk Museum voor Midden-Afrika in Tervuren (KMMA) heeft een belangrijke audiocollectie opgebouwd doorheen de twintigste eeuw. Het verzamelen van geluidsdocumenten startte in 1911 met voornamelijk opnames uit Centraal-Afrika, maar werd sindsdien uitgebreid met traditionele en volksmuziek uit de hele wereld op erg diverse – en soms intussen historische – geluidsdragers zoals wasrollen, Sonofil-draadopnamen, magneetband, vinylplaten, audiocassettes, DAT of cd. De collectie omvat ongeveer 5.000 uren aan muziekopnamen, waarvan het gedeelte 'Afrika' 2.550 uren omvat.

DEKKMMA (Digitalisatie van het Etnomusicologisch Klankarchief van het Koninklijk Museum voor Midden-Afrika) is het eerste grote digitaliseringsproject van het KMMA en beoogde de digitalisering van het hele klankarchief.

De audio werd gedigitaliseerd volgens standaarden voorgeschreven door de International Association of Sound and Audiovisual Archives (IASA)[1] en wordt bewaard op een RAID5E systeem: een multiprocessorstructuur waarbij de harde schijven beschermd zijn en gegevens opnieuw kunnen worden samengesteld in geval een schijf defect raakt.

Het kopiëren van de oudste geluidsdragers uit het archief, met name de wasrollen en Sonofil-draadopnamen gebeurde in samenwerking met het Phonogramm Archiv van het Ethnologisches Museum te Berlijn en de Vlaamse Radio en Televisie. De digitalisering van de

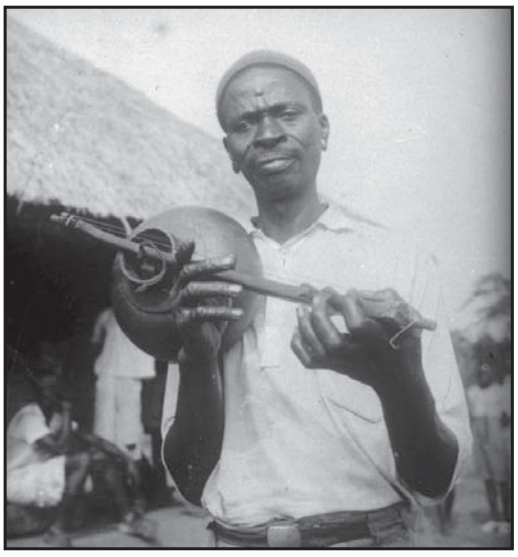
andere geluidsdragers werd in de klankstudio van het KMMA gerealiseerd.

Digitalisering op maat van de gebruiker

Naast de digitalisering van de klank, dienden ook de metadata (de beschrijving van de audio) gedigitaliseerd te worden. Daartoe werd een databank en invoerprogramma ontworpen. Voor het ontwerp werd vertrokken van de structuur en de semantiek van de aanwezige traditionele papieren steekkaarten. Het grote voordeel van een 'databank op maat' is dat het beter afgestemd is op specifieke noden: snellere invoer van en toegang tot de gegevens, controle op onderhoud en consistentie van de databank, mogelijkheid tot meertaligheid, de creatie van specifieke etnische beschrijvingsvelden en invoer van aanpassingen voor ondersteuning van musicologisch onderzoek.

Het ontwikkelen van een website[2] verhoogde de toegankelijkheid tot het archief. Niet alleen (fragmenten) audio en (alle) metadata zijn langs die weg beschikbaar, maar ook de nodige contextuele informatie om het archiefmateriaal te voorzien van een cultureel kader.

Digitalisering leidt tot een verhoogde toegankelijkheid tot muziek voor de gebruiker, die snel en eenvoudig zijn keuze moet kunnen terugvinden. Daartoe kunnen flexibele bevraging en Music Information Retrieval (MIR) een oplossing bieden. Een flexibel zoeksysteem impliceert dat niet enkel naar resultaten gezocht wordt die exact beantwoorden aan de zoekcriteria, maar ook naar resultaten die bij benadering beantwoorden aan de vraagstelling, wat in het bijzonder nuttig is indien er geen enkel resultaat exact voldoet aan alle criteria. Bovendien moeten niet alle criteria scherp gedefinieerd zijn, maar is er bij het opgeven van de zoekcriteria enige flexibiliteit



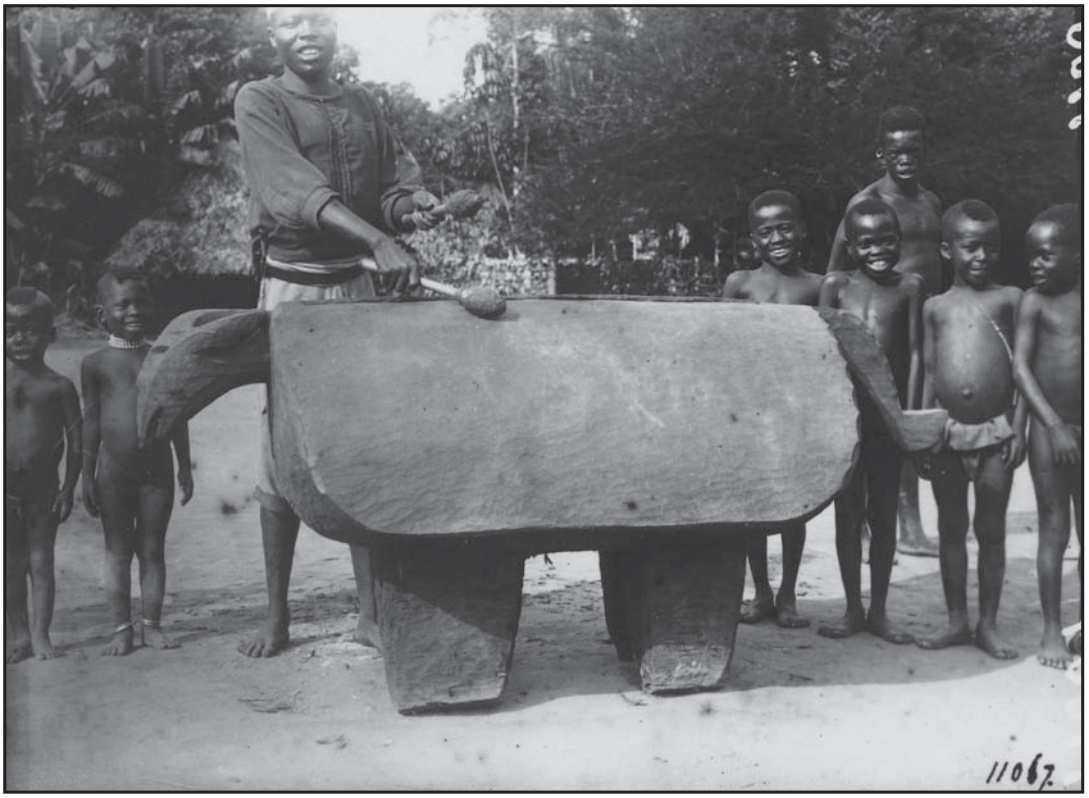
▶ De staafciter behoort tot de groep van de chordofonen: de instrumenten waarbij één of meerdere snaren tot trilling worden gebracht en zo klank produceren. De staafciter bestaat uit één snaar, opgespannen op een snaarhouder of hals die op een ronde klankkast, vaak een uitgeholde halve kalebas, gemonteerd is. De klankkast is vaak met geometrische of dierlijke figuren versierd. De staafciter begeleidde zang en dans bij verschillende volkeren in o.a. Congo, Kameroen, Rwanda en Zuid-Afrika.

▷ © Koninklijk Museum voor Midden-Afrika, Tervuren

toegestaan (bijvoorbeeld dichter aanleunend bij de natuurlijke taal).

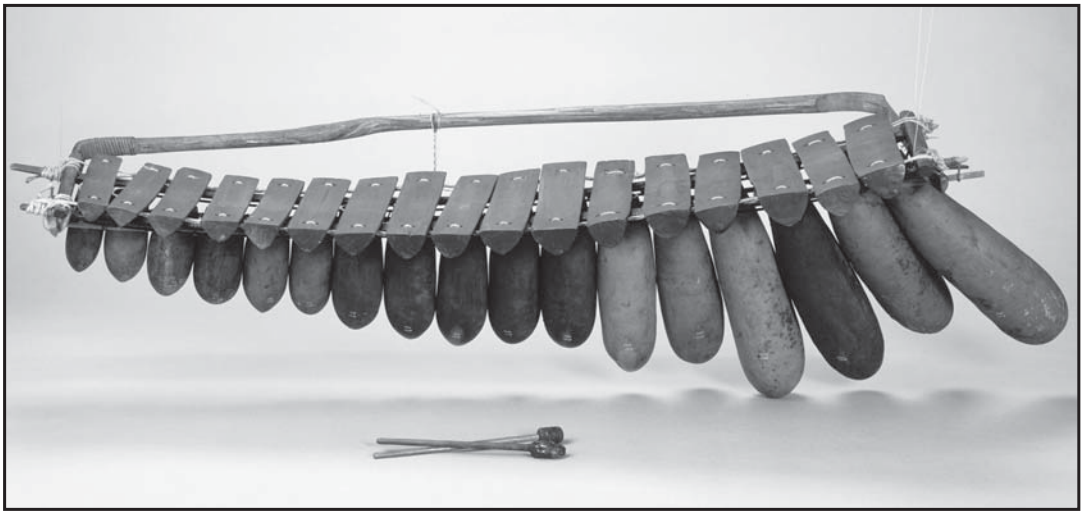
Voor het KMMA werd een methode ontwikkeld waarmee je het archief kan doorzoeken op basis van instrumentarium. Hierbij wordt er in het klankarchief gezocht naar opnames met een *gelijkaardig* klinkend instrumentarium, rekening houdend met het feit dat sommige instrumenten sterk op elkaar gelijken.

Bij het onderhoud van een digitaal klankarchief dient enige omzichtigheid geboden. Summier en in een niet-exhaustieve opsomming, dient nauwlettend gelet op 1) de duurzaamheid van de digitale drager (zogenaamde 'digitale broosheid'), 2) de connectie tussen digitale metadata en audio, 3) de bewaring van de beschrijving van de keuzes van het (doorgevoerde of geplande) digitaliseringsproces (zogenaamde *preservation*



▶ Door zijn ontwerp in de vorm van een dier (meestal een buffel of antiloop) is de zoömorfe spleettrommel een opmerkelijk instrument binnen het Afrikaanse instrumentarium. Dergelijke trommels bestaan uit een uitgehold stuk hout met een gleuf als klankgat. Het uithollen gebeurt aan weerszijden van de gleuf in verschillende diktes, zodat twee toonhoogtes gespeeld kunnen worden op het instrument. Als seininstrument en als ritmisch begeleidingsinstrument van dansen had de zoömorfe spleettrommel een belangrijke maatschappelijke functie. De grote zoömorfe trommen zijn nagenoeg uitsluitend afkomstig uit de Ubangi-streek, die zich uitstrekt over Congo en de Centraal-Afrikaanse Republiek.

▷



▶ De *madimba* is een van de meest complexe instrumenten die het Afrikaanse instrumentarium kent. Het bezitten van een dergelijke xylofoon was het voorrecht van personen met aanzien binnen de lokale gemeenschap, waaronder de lokale chefs. Door de grote technische vaardigheid die dit instrument vereist, staan ook de bespelers van de *madimba* in hoog aanzien. Vaak wordt het instrument in 'paren' ingezet, waarbij het ene instrument de melodie speelt en het andere instrument een ostinaat begeleidend motief uitvoert. De *madimba* kent zijn grootste verspreiding in de vroegere Congolese provincies Kasaï en Katanga.

▷ © Koninklijk Museum voor Midden-Afrika, Tervuren

of the preservation), 4) de leesbaarheid van gekozen digitale formaten, 5) de uitwisselbaarheid van de gegevens van de databank (belang van exportmogelijkheden van gegevens via een gestandaardiseerd formaat, daar steeds meer databases van verschillende musea gekoppeld zullen worden tot een grote online databank). Hiermee is ook de uitwisselbaarheid met andere archieven op langere termijn verzekerd en wordt de ontsluiting van het eigen archief bevorderd.

Music Information Retrieval

Dankzij het MIR-onderzoek[3] wordt het mogelijk om naar muziek te zoeken op basis van muzikale inhouden. Met MIR wordt immers bedoeld: geavanceerde technologie om, via analyse van klank en partituur, muzikale kenmerken te ontginnen, zodanig dat specifieke muziekbeschrijvingen, inhoudsgebaseerde muziekzoeksystemen en vernieuwend musicologisch onderzoek ontwikkeld en gestimuleerd worden. MIR levert een methodische en alternatieve kijk op audio op en kan grosso modo op drie niveaus plaatsvinden:

1) Analyse van het fysieke klanksignaal, mogelijk door analyse van digitale audio. Immers, muziek in analoge vorm met zijn continue signaal en daardoor aan subjectievere normen onderworpen beschrijving, wordt na digitalisering discreet en daardoor meetbaar op microniveau, hetgeen

(beperkte) menselijke auditieve analyses sterk verrijkt, zowel kwantitatief als kwalitatief.

2) Analyse op symbolisch/semantisch niveau, waarbij de digitale partituur statistisch geanalyseerd wordt.
3) Analyse op contextueel niveau, bijvoorbeeld door *web-tagging*, een methode waarbij muziekbeschrijvingen aangeleverd worden door kwantitatieve verzamelingen van webdata.

Het merendeel van de MIR toepassingen richt zich op commerciële muziekgenres, met als gevolg dat de bestaande toepassingen zelden geschikte beschrijvingen aanleveren voor niet-commerciële muziek. Enerzijds is de software immers geoptimaliseerd voor specifieke genres, anderzijds vraagt etnische muziek om geheel andersoortige beschrijvingen van klankkenmerken die niet zomaar te vatten zijn in het klassieke muziekjargon. Het gevaar van een dergelijke gesloten keten bestaat er in dat mensen die zoeken naar muziek altijd terechtkomen bij Westerse commerciële muziek en dat niet-commerciële muziek minder snel teruggevonden wordt.

MIR-onderzoek kan geleverd worden op alle muzikale parameters zoals toonhoogte, toonladders, timbre, tempo en ritmische patronen. Voor het DEKKMMA project werden voor elk van de duizenden geluidsopnames van het KMMA-archief muzikale gegevens automatisch geëxtraheerd en vervolgens verwerkt tot concrete datasets en een grafische weergave. De unieke precisie van de muzikale data toont kenmerken van etnische muziek die vroeger erg moeilijk te meten waren. De databank kan nu doorzocht worden op een manier die voorheen niet mogelijk was, namelijk op basis van muzikale inhouden. Het broze etnisch-culturele erfgoed kan op die manier meer bekendheid en toegankelijkheid krijgen. MIR opent bovendien enorme mogelijkheden voor innovatief musicologisch onderzoek.

Eindnoten

1. www.iasa-web.org.
2. http://music.africamuseum.be.
3. www.ismir.net.