5th International Workshop on Folk Music Analysis

June 10-11-12, 2015

University Pierre and Marie Curie

Paris, France

http://fma2015.sciencesconf.org/

Institut Jean le Rond d'Alembert
CNRS UMR 7190
Lutheries-Acoustique-Musique

9 791095 209003

Edited by Association Dirac

The 5th International Workshop on Folk Music Analysis (http://fma2015.sciencesconf.org) was organized by the LAM Team (Lutheries – Acoustique – Musique) from the Institute D'Alembert CNRS UMR 7190.
This 2015 session was hosted by the University Pierre and Marie Curie (www.upmc.fr), Paris, France, on 10, 11 and 12 June 2015.

The FMA workshop is intended to provide a platform for new research and discussion on the topics:
•Computational (ethno)musicology, Retrieval systems for non-western and folk musics
•New methods for music transcription
•Statistical approaches to music
•Empirical approaches to music
•Folk music classification systems
•Models of oral transmission of music
•Cognitive modelling of music
•Aesthetics and related philosophical issues
•Methodological issues
•Representational issues and models
•Audio and symbolic representations
•Formal and computational music analysis
Specific research topics, fields of study, and methodological approaches have been left open intentionally to encourage interdisciplinary exchange.

We invited 3 keynote speakers:
Prof. François PICARD, IReMus UMR 8223, Paris-Sorbonne, FRANCE
Dr Julien PINQUIER, Team Structuration, Analyse et MOdélisation de documents Vidéo et Audio, IRIT, Toulouse, FRANCE
Prof. Gaël RICHARD, Audio, Acoustics and Waves Group, Télécom ParisTech, Paris, FRANCE

Thanks to the scientific committee (*alphabetic order*): ADAM Olivier, ASSAYAG Gérard, BENETOS Emmanouil, BONINI BARALDI Filippo, BURGOYNE Ashley, CAMBOUROPOULOS Emilios, CAZAU Dorian, CHEMILLIER Marc, CLER Jérôme, CONKLIN Darrell, DEBOVE Julien, DOVAL Boris, ESTIVAL Jean-Pierre, FILLON Thomas, GLOTIN Hervé, GOMEZ Emilia, GOMEZ Paco, HOLZAPFEL Andre, LARTILLOT Olivier, PELLERIN Guillaume, PICARD François, PINQUIER Julien, SIMONNOT Joséphine, VAN KRANENBURG Peter, and WEYDE Tillman

Special thanks to (*alphabetic order*) BIDOUZE Marianne, CAZAU Dorian, MATHIEU Justine, lORTHIOIS Chloé, SOUFFLET Clément, the LAM Team, the ADAM family, and the Association DIRAC (France)
Olivier ADAM

And of course, a big thank you to *you* for taking part in this 2015 FMA workshop !

Olivier ADAM and Dorian CAZAU

PROGRAM OF THE FMA 2015 WORKSHOP

June, 10

Welcome to the participants	
8:00 – 9:00am	Registration
9:00 – 9:05am	Welcome from Prof Paul Indelicato, Vice-President of the Scientific Council, University Pierre et Marie Curie, France
9:05 – 9:10am	Welcome from Prof Stéphane Zaleski, Director of the Institut Jean Le Rond d'Alembert, University Pierre et Marie Curie, France
9:10 – 9:15am	Welcome from Hugues Genevois, Director of the Lutheries Acoustics Music Team, D'Alembert, UPMC, France
9:15 – 9:20am	Welcome from Véronique Atger, Director of the Research Department, Sorbonne Universities, France
9:20 – 9:25am	Welcome from Olivier Adam, FMA Organisor, University Pierre et Marie Curie, France
Invited Speaker #1	
9:30 – 10:30am	Prof François Picard, IReMus UMR 8223, Paris-Sorbonne, France *Monika and the Diagram of melodic analysis. A collectively elaborated tool for computer-aided analysis of pentatonic, modal, and tonal music*
10:30 – 11:00am	*Coffee break*
Oral Session #1	**Topic: Contextual analysis Chairman: Peter van Kranenburg**
11:00 – 11:30am	Yosr Bouhali, Jean-Dominique Polack: *The propulsion of the takht in the new acoustic frame: cycle of evolution and acoustic indications*
11:30 – 12:00pm	Varun deCastro-Arrazola, Peter van Kranenburg, Berit Janssen: *Computational textsetting analysis of Dutch folk songs*
12:00 – 12:30pm	Ana Lucia Fontenele: *Analysis of Brazilian popular music: the Pixinguinha arrangements for the orchestra's of the radio program O Pessoal da Velha Guarda*
12:30 – 1:00pm	Polina Proutskova: *Approaching vocal production in world's music cultures - a mixed methods study based on the physiology of singing*
Poster session #1	
1:00 – 2:00pm	Poster session during the *Lunch*
Oral Session #2	**Topic: Classification Chairman: Olivier Adam**
2:00 – 2:30pm	Darrell Conklin, Kerstin Neubarth, Stéphanie Weisser: *Contrast pattern mining of Ethiopian bagana songs*
2:30 – 3:00pm	Kerstin Neubarth: *Densmore Revisited: Contrast Data Mining of Native American Music*
3:00 – 3:30pm	Pál Richter: *Folk Music Classifications in Hungarian Ethnomusicology*
3:30 – 4:00pm	Islah Ali-MacLachlan, Munevver Kokuer, Peter Jancovic, Cham Athwal: *Towards the identification of irish traditional flute players from commercial recordings*
Poster session #2	
4:00 – 4:30pm	Poster session during the *Coffee break*
Oral Session #3	**Topic: Quantitative analysis Chairman: Barış Bozkurt**
4:30 – 5:00pm	Frank Scherbaum, Wolfgang Loos, Frank Kane, Daniel Vollmer: *Body vibrations as source of information for the analysis of polyphonic vocal music*
5:00 – 5:30pm	Moreno Andreatta, Mattia Bergomi: *Math'n pop versus math'n folk? a computational (ethno)musicological approach*
5:30 – 6:00pm	Bilge Mirac Atici, Baris Bozkurt, Sertan Senturk: *A Culture-Specific Analysis Software for Makam Music Traditions*
IceBreaking evening	
7:30 – 9:30pm	Invitation to the IceBreaking evening at *UPMC, Jussieu, Tour Zamansky, 24^{e} floor*

June, 11

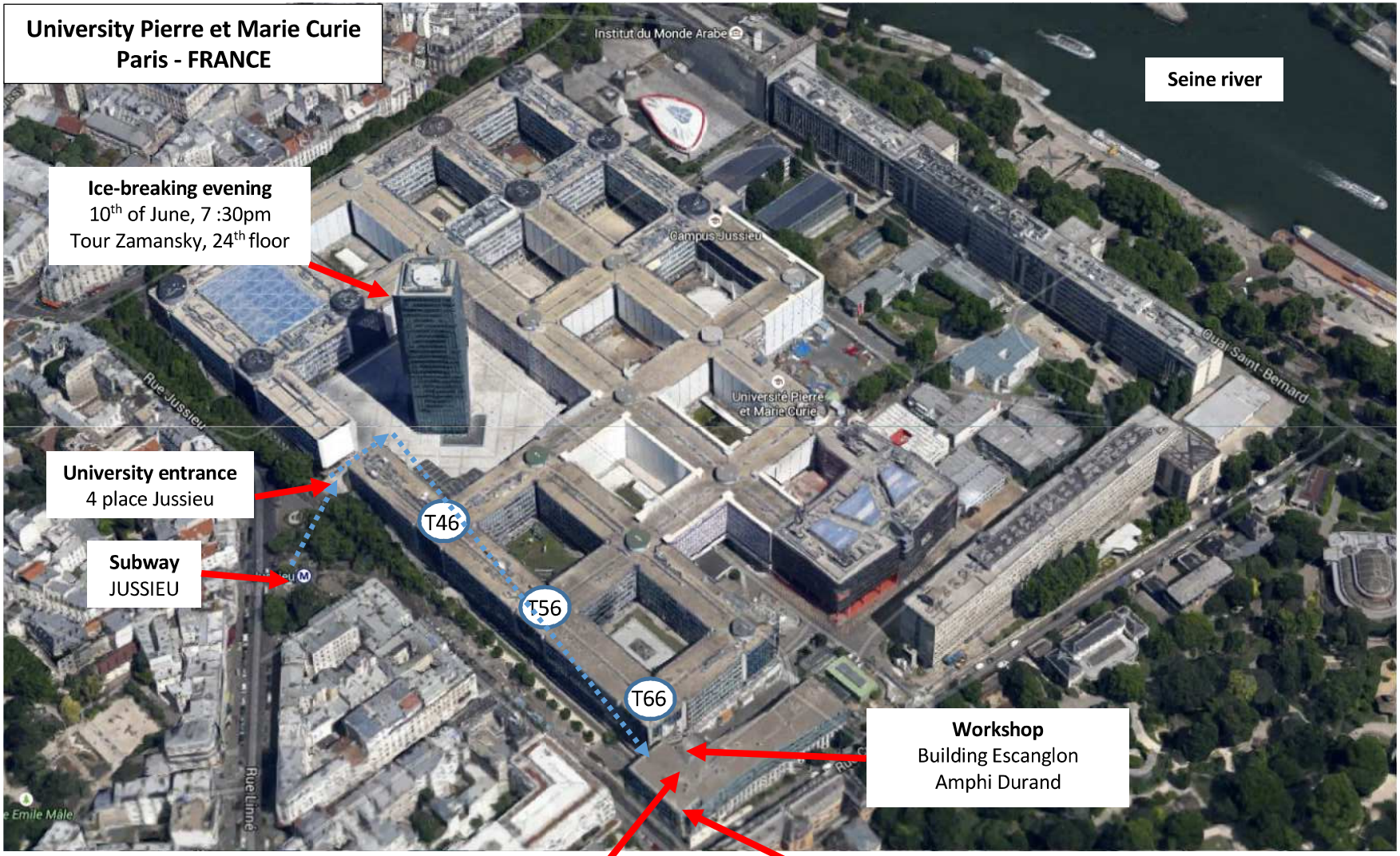
Invited Speaker #2	
9:00 – 10:00am	Prof Gaël Richard, AAWG, Télécom ParisTech, France: *Melody Extraction from music signals: "blind" and "informed" approaches*
Poster session #3	
10:00 – 10:30am	Poster session during the *Coffee break*
Oral Session #4	**Topic: Music transcription and pitch analysis** **Chairman: Tillman Weyde**
10:30 – 11:00am	Dorian Cazau, Olivier Adam: *Automatic Music Transcription of the Marovany (Madagascar) zither repertoires based on prior knowledge from musical acoustics*
11:00 – 11:30am	Paolo Bravi, Cécile Delétré, Marco Lutzu, Emmanuelle Olivier, François Picard, Alice Tacaille: *The Hukwe bow song of the symposium on transcription (Middletown, 1963) fifty years later: new pespectives, methodologies and analyses*
11:30 – 12:00pm	Hasan Sercan Atli, Baris Bozkurt, Sertan Sentürk: *A Method for Tonic Frequency Identification of Turkish Makam Music Recordings*
12:00 – 12:30pm	Samer Abdallah, Aquiles Alencar-Brayner, Emmanouil Benetos, Stephen Cottrell, Jason Dykes, Nicolas Gold, Alexander Kachkaev, Mahendra Mahey, Dan Tidhar, Adam Tovell, T. Weyde, D. Wolff: *Automatic transcription and pitch analysis of the British Library World & Traditional Music Collections*
Poster session #4	Authors have 3min to orally pitch their poster
12:30 – 12:33pm	Luwei Yang, Mi Tian, Elaine Chew: *Vibrato Characteristics and Frequency Histogram Envelopes in Beijing Opera Singing*
12:33 – 12:36pm	Frank Scherbaum, Simha Arom, Frank Kane : *On the feasibility of Markov model based analysis of Georgian vocal polyphonic music*
12:36 – 12:39pm	Burak Uyar, Baris Bozkurt: *An Interactive Rhythm Training Tool for Usuls of Turkish Makam Music*
12:39 – 12:42pm	Wang Yuancheng, Dorian Cazau, Olivier Adam: *Characterizing complex note temporal profiles with the PLCA-HMM method*
12:42 – 12:45pm	Johan Loeckx: *Fluid Construction Grammar: a new computational paradigm for studying music and meaning*
12:45 – 2:00pm	*Lunch*
Oral Session #5	**Topic: Melody analysis and similarity** **Chairman: Andre Holzapfel**
2:00 – 2:30pm	Izaro Goienetxea, Darrell Conklin: *Transformation of a bertso melody with coherence*
2:30 – 3:00pm	Andre Holzapfel: *Melodic key phrases in traditional Cretan dance tunes*
3:00 – 3:30pm	Chris Walshaw: *Multilevel melodic matching*
3:30 – 4:00pm	Jose-Miguel Díaz-Báñez, Nadine Kroher, Juan Carlos Rizo: *Efficient algorithms for melodic similarity in flamenco singing*
4:00 – 4:30pm	Anas Ghrab: *A simple method for a melodic classification*
Poster session #5	
4:30 – 5:00pm	Poster session during the *Coffee break*
Special Session	**Topic: Automatic music improvisation**
5:00 – 6:00pm	Prof Marc Chemillier, CAMS-EHESS, Paris, France : *The ImproteK project: presentation and demonstration*
Poster session #6	
6:00 – 6:30pm	Poster session
Concert	
8:00 – 10:00pm	**Kilema**, concert of Malagasy music Jussieu, Building Escanglon, cave floor

June, 12

Invited Speaker	
9:00 – 10:00am	Dr Julien Pinquier, IRIT, Toulouse, France *DIADEMS: Description, Indexing, Accessibility to ethnomusicologist and sound documents*
Poster session #7	
10:00 – 10:30am	Poster session during the *Coffee break*
Oral Session #6	**Topic: Data Mining for Ethnomusicological Dataset** **Chairman: Julien Pinquier**
10:30 – 10:55am	Guillaume Pellerin, Joséphine Simonnot, Thomas Fillon: *Web analysis tools and ethnomusicology*
10:55 – 11:20am	Jean-Luc Rouas, Dominique Fourer: *Classifying instruments with timbral features : application to ethnomusicological recordings*
11:20 – 11:45am	Marwa Thlithi, Claude Barras, Julien Pinquier, Thomas Pellegrini: *Singer Diarization: Application to ethnomusicological recordings*
11:45 – 12:10pm	Lionel Feugère, Boris Doval, Marie-France Mifune: *Using pitch features for the characterization of intermediate vocal productions*
12:10 – 12:35pm	Armand Leroi, Matthias Mauch, Pat Savage, Emmanouil Benetos, Juan Bello, Maria Panteli, Joren Six, Tillman Weyde: *The deep history of music project*
12:35 – 1pm	Apollinaire Anakesa: *Towards a New Hearing of Music from the World of Orality: the case of African and Caribbean musical cultures*

Posters

Marcelo Rodriguez, Dimitrios Bountouridis, Anja Volk: *A novel music segmentation interface and the Jazz theme dataset*
Luwei Yang, Mi Tian, Elaine Chew: *Vibrato Characteristics and Frequency Histogram Envelopes in Beijing Opera Singing*
Frank Scherbaum, Simha Arom, Frank Kane: *On the feasibility of Markov model based analysis of Georgian vocal polyphonic music*
Burak Uyar, Baris Bozkurt: *An Interactive Rhythm Training Tool for Usuls of Turkish Makam Music*
Johan Loeckx: *Fluid Construction Grammar: a new computational paradigm for studying music and meaning*
Wang Yuancheng, Dorian Cazau, Olivier Adam: *characterizing complex note temporal profiles with the PLCA-HMM method*



Rmq: to access the Building Escanglon, you have to walk inside the University

- Use the main entrance of the University
- Make a right to the Towers 46, 56 and 66

CONTENTS

INVITED SPEAKERS

(*alphabetic order*)

Prof Marc Chemillier, CAMS-EHESS, Paris, France

The ImproteK project.
Presentation and demonstration

ImproteK is a musician-machine interaction system that learns in real time from human performers and generates improvisations in the style of these performers. It produces new musical phrases by recombining sequences played by a real musician. As such ImproteK belongs to the OMax family and derives from earlier works dedicated to the OMax software done at IRCAM which were based on sequence modeling and statistical learning. The new issue addressed by ImproteK is to take into account an underlying metrical structure in the music recombination process. The ImproteK environment proposes through a retrieval/recombination process of musical motives to explore the creative potential of traditional musicians, which they exploit when developing their own repertoires from the base original motives specific to their cultural patrimony. This meeting between folk music and computer-based interactive environments leaves much room for experiments on musical creativity. It would also contribute in better representing folk music in contemporary human-machine performances.

Prof François Picard, IReMus UMR 8223, Paris-Sorbonne, France

Monika and the Diagram of Melodic Analysis.
A Collectively Elaborated Tool for Computer-Aided Analysis
of Pentatonic, Modal, and Tonal Music

The diagram of melodic analysis is a tool which for a given melody expresses a dimension of melodic movement, a weighted hierarchy of notes, the “relative melodic importance of the notes”, the "melodic weight", a modal pattern. It is evocative of a dimension of music which cannot expressed only through the list of the notes used and the finalis. The diagram presented here is the product of a collective elaboration above a century, from Erich von Hornbostel to Nicolas Meeùs and Vincent Boucheau through Rulan Chao Pian, Monika Stern, and Alice.

Dr Julien Pinquier, IRIT, Université de Toulouse, France

Diadems:
Description, Indexing,
Accessibility to ethnomusicologist
and Sound Documents

The DIADEMS project should enable technology providers to offer content indexing tools through the emergence of acoustic attributes and adequate modeling. For the musicologist and the ethnolinguist the objective goes beyond indexing and annotation opportunities offered him. The goal is to have the relevant tools to assist in information extraction.

Prof. Gaël Richard, Signal and Image Processing Department, Télécom ParisTech, France

Melody Extraction From Music Signals:
“Blind” and “Informed” Approaches

Melody extraction algorithms aim at extracting a sequence of frequency values corresponding to the pitch of the dominant melody from polyphonic musical signals. The melody line can be extracted using the downmix audio signal only (the so-called "blind" approach) or using additional side-informations such as the musical score or a hummed copy of the melody (the so-called "informed" approach).
The talk will provide an overview of both trends for main melody extraction from polyphonic audio signals.

AUTOMATIC TRANSCRIPTION AND PITCH ANALYSIS OF THE BRITISH LIBRARY WORLD & TRADITIONAL MUSIC COLLECTIONS

Samer Abdallah[1], Aquiles Alencar-Brayner[2], Emmanouil Benetos[5], Stephen Cottrell[4] Jason Dykes[3], Nicolas Gold[1], Alexander Kachkaev[3], Mahendra Mahey[2], Dan Tidhar[4], Adam Tovell[2], Tillman Weyde[3], Daniel Wolff[3]

[1] Department of Computer Science, University College London, [2] The British Library,
[3] Department of Computer Science, City University London,
[4] Department of Music, City University London
[5] Centre for Digital Music, Queen Mary University of London
dml-owner@city.ac.uk - http://dml.city.ac.uk

1. INTRODUCTION

Music research, particularly in fields like systematic musicology, ethnomusicology, or music psychology has developed as "data oriented empirical research" (Parncutt, 2007), which benefits from the development of computing methods and infrastructure. In ethnomusicology there has been a recent growing interest in computational methods and their application to audio data collections (Gómez et al., 2013; Canazza et al., 2010), and in the degree to which such methods may reveal insights into musical practice which may not be evident from the participant-observation paradigms that have otherwise characterised the discipline. For technological and legal reasons, research in this area was previously limited to small datasets, but this is changing in part due to the contribution of research projects such as CompMusic [1] and cultural preservation projects such as Europeana Sounds [2].

The authors of this paper collaborate in the UK AHRC-funded project *Digital Music Lab - Analysing Big Music Data (DML)* [3] to develop methods and technologies to support the use of Big Data in musicology (Weyde et al., 2014). As part of the project, we have developed a software/hardware infrastructure for exploring and analysing large-scale audio collections, aiming to assist research in systematic and empirical (ethno)musicology.

A major partner in the DML project is The British Library (BL), which holds several million audio recordings in its Sound Archive, spanning oral history interviews, environmental, and natural sounds, as well as over 3 million recordings from classical, popular, world and traditional music (of which approximately 10% are digitized). A portion of these audio recordings (approx. 60,000) are currently available for online streaming [4]. Through the DML project, a computing server was installed on-site at the BL, enabling storage and analysis for a collection of over 29,000 recordings from its 'World & Traditional Music' corpus.

In this paper we present the collection of World & Traditional music that was curated and analysed as part of the DML project; we will also present methods for automatic transcription and pitch analysis that were applied to recordings from that collection and were used as a basis for creating an integrated tool/interface enabling musicological enquiries and research in large music collections.

2. COLLECTION

The BL Sound Archive holds one of the world's largest collections of recordings variously described as traditional, folk or 'world' music. The dataset drawn from this collection for the DML project consists of 29,198 audio recordings. It covers a large collection (8,000) of English, Irish, and Scottish folk songs; 1300 recordings from Oceania; 12,000 recordings from Africa (covering West and South Africa, as well as large collections from Uganda and Sudan); over 6,000 recordings from Asia (mostly from Nepal, India, and Pakistan); 1100 recordings from the Middle East; and a small collection of 47 recordings from the Americas (comprising music of indigenous people from Colombia). It also contains collections of wax cylinders recorded by pioneering fieldworkers, as well as more recent recordings made in a range of formats as part of ethnographic research. Recording dates span from 1898 (from the ethnographic wax cylinders collection) up to 2014. It is worth noting that several of these recordings contain segments of speech as well as music, or even overlapped speech and music. Information on the five largest collections can be seen in Table 1.

The recordings are also accompanied with rich metadata in METS/XML format. Information present in the metadata includes: title, collection ID, description, performer, recording engineer, recording date and temporal information (e.g. Easter), language, geographic information, as well as audio file information (duration, sampling rate, resolution, URL for publicly available recordings).

Authors in alphabetical order. This work was supported by the UK AHRC-funded project 'Digital Music Lab - Analysing Big Music Data', grant no. AH/L01016X/1 and the UK AHRC funded project 'An Integrated Audio-Symbolic Model of Music Similarity', grant no. AH/M002454/1. EB is supported by a Royal Academy of Engineering Research Fellowship, grant no. RF/128.

[1] http://compmusic.upf.edu/
[2] http://www.europeanasounds.eu/
[3] http://dml.city.ac.uk
[4] http://sounds.bl.uk/

Title	# recordings	Dates
Bob & Jacqueline Patten English Folk Music Collection	6333	1953-2002
John Howson English, Irish & Scottish Folk Music Collection	3498	1930-1999
Reg Hall English, Irish & Scottish Folk Music & Customs Collection	3195	1949-1996
Klaus Wachsmann Uganda Collection	1538	1949-1954
Peter Cooke Uganda Collection	1277	1964-1997

Table 1: The five largest collections from the BL World & Traditional Music dataset used for the DML project.

3. AUTOMATIC TRANSCRIPTION

As a first step towards musicological analysis for large audio collections, we use automatic music transcription (AMT) technology to convert a recording into machine-readable music notation (Klapuri & Davy, 2006). The vast majority of research in AMT technology is however limited to Western/Eurogenetic music, where an audio recording is converted into a MIDI-like representation.

For this work we used the model of Benetos & Dixon (2012), which can support the estimation of multiple pitches (along with onsets, offsets, and velocities) in a scale equal to the resolution of the input time/frequency representation (in our case, 20 cent resolution given as input a log-frequency spectrogram of 60 bins/octave), and ranked first in the MIREX 2013 evaluations for Multiple-F0 estimation and Note Tracking [5]. This method is also publicly available as a VAMP plugin [6], which can be used in conjunction with software such as Sonic Visualiser [7].

The output of the aforementioned method is a probability distribution of pitches in 20 cent resolution over time: $P(f,t)$ (f corresponds to pitch and t to the time index). This is post-processed and converted into a binary representation of note events, with a corresponding onset time, offset time, pitch, and velocity value. Apart from high-level musicological applications the resulting transcription can serve as a way to store or visualise the content of a music recording (an example can be shown in Fig. 1). The resulting transcription files, along with other low-level features, can be downloaded for individual files through the semantic web server of the DML project [8].

4. PITCH HISTOGRAMS

Information extracted from the automatic transcriptions can be used to generate information on the pitch content of a recording, or a collection of recordings. Given a transcription $P(f,t)$, a pitch histogram for that recording can be generated by: $P(f) = \sum_t P(f,t)$. The above process can be extended for any number of recordings, for visualising the pitch content of collections. As part of the DML interface for browsing/interacting with large music collections (Kachkaev et al., 2015), pitch histograms for audio collections can be computed on-demand. The DML visualisation

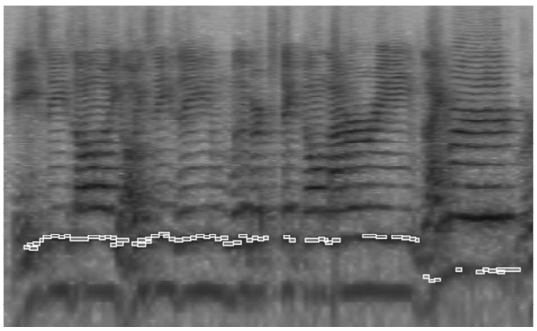


Figure 1: The spectrogram for a segment of 'The Blackbird' from the Bob & Jacqueline Patten Collection. Overlayed as white boxes are detected pitches.

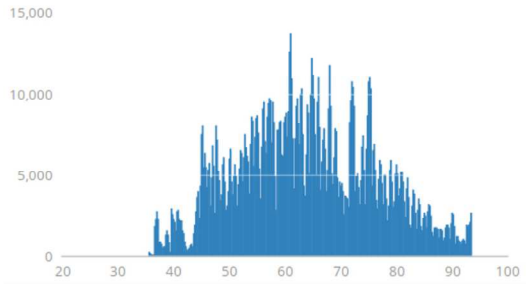


Figure 2: The aggregated pitch histogram for 3,022 Ugandan recordings, as shown in the DML VIS interface.

interface is available online [9]. An example pitch histogram taken from that interface is shown in Fig. 2.

This allows the identification of the most frequently encountered pitches within a given collection, which in turn may suggest pitch hierarchies prevailing within the music culture and, potentially, commonly used scales. Cross-referencing with ethnographically-grounded research is likely to be necessary to confirm the cultural validity of the results. Nevertheless, the method allows large-scale examination of particular music cultures, comparison between collections to observe similar patterns of pitch distribution and, possibly, the automatic identification of otherwise unattributed recordings according to the pitch distribution found in them.

5. DISCUSSION

In this paper, we presented work carried out by the DML project on analysing over 29,000 recordings from the BL World & Traditional Music collection. This involved producing automatic transcriptions at a high pitch resolution and creating pitch histograms both for individual recordings and collections. These are supported by a semantic web server for downloading features for individual recordings and a visualisation interface for browsing through collections.

Open questions remain about the use of additional high-level features for describing folk and traditional music recordings, given culturally-specific aspects such as modes, structure, and instrumentation. We believe that pitch anal-

[5] http://www.music-ir.org/mirex/wiki/MIREX_HOME
[6] https://code.soundsoftware.ac.uk/projects/silvet/files
[7] http://sonicvisualiser.org/
[8] http://mirg.city.ac.uk/cp/
[9] http://dml.city.ac.uk/vis/

ysis at a high-frequency resolution is the first step towards that goal. Another issue relates to the integration of recordings from copyright-restricted collections; the recent UK Intellectual Property Bill has enabled researchers to use copyright-restricted data for research purposes, but computations still need to be done on-site. We believe this approach will enable musicians and musicologists unprecedented access to recordings, overcoming such restrictions. Finally, the problem of segmenting recordings into speech and music parts is currently being investigated through the MuSpeak project [10].

6. REFERENCES

Benetos, E. & Dixon, S. (2012). A shift-invariant latent variable model for automatic music transcription. *Computer Music Journal*, *36*(4), 81–94.

Canazza, S., Camurri, A., & Fujinaga, I. (2010). Special section: Ethnic music audio documents: From the preservation to the fruition. *Signal Processing*, *90*(4), 977–1334.

Gómez, E., Herrera, P., & Gómez-Martin, F. e. (2013). Special issue: Computational ethnomusicology. *Journal of New Music Research*, *42*(2).

Kachkaev, A., Dykes, J., Abdallah, S., Barthet, M., Benetos, E., Cottrell, S., Dixon, S., Gold, N., Hargreaves, S., Tidhar, D., Wolff, D., & Weyde, T. (2015). Small multiples for big data a framework for comparison using open web technologies with music collections. In *IEEE Information Visualization Conference 2015, submitted*.

Klapuri, A. & Davy, M. (Eds.). (2006). *Signal Processing Methods for Music Transcription*. New York: Springer-Verlag.

Parncutt, R. (2007). Systematic musicology and the history and future of western musical scholarship. *Journal of Interdisciplinary Music Studies*, *1*, 1–32.

Weyde, T., Cottrell, S., Dykes, J., Benetos, E., Wolff, D., Tidhar, D., Gold, N., Abdallah, S., Plumbley, M., Dixon, S., Barthet, M., Mahey, M., Tovell, A., & Alencar-Brayner, A. (2014). Big data for musicology. In *Digital Libraries for Musicology Workshop*, (pp. 85–87)., London, UK.

[10] http://mirg.city.ac.uk/muspeak

TOWARDS THE IDENTIFICATION OF IRISH TRADITIONAL FLUTE PLAYERS FROM COMMERCIAL RECORDINGS

Islah Ali-MacLachlan, Münevver Köküer, Cham Athwal
School of Digital Media Technology,
Birmingham City University, UK
{islah.ali-maclachlan, munevver.kokuer,
cham.athwal}@bcu.ac.uk, m.kokuer@bham.ac.uk

Peter Jančovič
School of Electronic, Electrical
& Systems Engineering
University of Birmingham, UK
p.jancovic@bham.ac.uk

ABSTRACT

This paper explores whether distinct spectral differences exist between professional flute players of Irish traditional music and between their playing of different notes and whether we could identify players based on their spectral differences in commercial recordings. The audio signal is represented using short-term magnitudes of the first few harmonic frequencies. Player identification is performed by employing the Gaussian classifier, where each player is characterised using a single multivariate Gaussian distribution with full covariance matrix. Experimental evaluations are performed using audio recordings from five professional flute players. Only the sustained sections of notes, which were manually identified based on note onsets, are used. The identification of players is explored for each note separately and using various number of note instances.

1. INTRODUCTION

Irish Traditional Music is a melodic form of instrumental music that was originally used to accompany dancing. It has, along with many forms of folk music, undergone a revival and is enduringly popular for players and listeners alike. Informal 'sessions', whilst a relatively modern phenomenon, allow flute players to gather with players of other traditional and modern melody and rhythm instruments including fiddle (violin), tin whistle, mandolin, accordion, guitar and bouzouki (Williams, 2010). Whilst new music is constantly being added to the tradition, the melodies collected by O'Neill (1998) and first published in 1907 still form a central corpus.

The most popular style of flute used by traditional players is the wooden concert D flute. It evolved from the simple system unkeyed flute and is available in both unkeyed (diatonic) and keyed (chromatic) styles. It continued to be used by traditional players long after Boehm's metal concert flute became the norm for orchestral players (Hamilton, 1990). One of the main reasons for this, as noted by Breathnach (1996) is that ornaments such as sliding into notes and rolling, a very fast inter-note articulation, are not possible with Boehm's key design.

Playing a flute requires control over breath pressure, lip position and lip aperture (Coltman, 1968). Changes in these parameters will result in timbral differences, a contributing factor to tonal differences between players. Timbre is the quality of a musical note that distinguishes it from other notes of the same pitch and loudness. Schouten (1968) proposed five acoustic parameters to describe timbre: Range between tonal and noiselike character, spectral envelope, changes in spectral envelope and fundamental frequency over time, sound onset and time envelope.

Erickson (1975) agrees that these dimensions are excellent for perceptual analysis of music. Handel (1995) argues that there are many stable and time-varying acoustic properties but no single property fully defines timbre. Burred et al. (2010) found that temporal envelope is more valuable when distinguishing between sustained and decaying instruments. In this study we do not analyse notes based on temporal attributes.

According to Jensen (1999) spectral envelope is often enough to identify a sound. Timoney et al. (2004) found that in spectral magnitude plots of tin whistles, a closely related instrument to the flute, there are only a small number of harmonics compared to other instruments. Chudy & Dixon (2010) presented a cello player recognition system employing timbre features. They used controlled studio recordings of 6 cellists playing the same excerpt on two different cello instruments.

Utilising spectral analysis, our earlier study showed clear timbral differences between players with varying levels of experience, each playing single G_4 notes as part of a scale with different models of wooden flute (Ali-MacLachlan et al., 2013). We found that magnitudes of F_1, F_2 and F_4 varied the most between individual players. This work followed on from Widholm et al. (2001) who identified timbral differences between individual classical flautists playing Boehm system flutes manufactured from a range of metals. Both studies showed that individual players have significantly more effect on timbre than changes in material or pattern.

To build on our earlier study (Ali-MacLachlan et al., 2013), we explore whether distinct spectral differences exist between professional players and between their playing of different notes and whether we could identify professional players based on their spectral differences in commercial recordings. We analyse the playing of five players, all at a professional level, playing reels, jigs, hornpipes and polkas. These are some of the most common forms of Irish traditional dance tunes. We investigate timbre-identity of players across the notes that have the highest usage in the corpus, specifically, D_4, E_4, $F\#_4$, G_4, A_4, B_4, D_5, E_5 and $F\#_5$. The analysis is performed using the sustained sections

of the notes. The short-term magnitudes of the first few harmonics are used as timbre descriptors. Player identification is performed using Gaussian classifier, where each player is modelled using a single multivariate Gaussian distribution with full covariance matrix. Four-fold cross-validation procedure is used to obtain player identification results.

2. AUTOMATIC PLAYER IDENTIFICATION

2.1 Acoustic analysis

The volume differences between the recordings are normalised to ensure that each recording has the same average energy. The signal is then segmented into short overlapping signal frames. Each signal frame is multiplied by the Hamming window function. The windowed frames are then zero padded, the Fourier transform is applied, and the absolute value taken to provide short-term magnitude spectrum. The magnitude values are compressed by applying logarithm. We used here the short-term magnitudes of the first four harmonics as timbre descriptors. Harmonics were identified semi-automatically based on the annotation, i.e., the harmonics were located by finding peaks around the multiples of the note frequency provided by the label file. In this initial study we explore the variations in timbre between players at the note level. All notes were extracted from the entire recordings and the analysis was performed for each type of note separately. Notes corresponding to ornaments are discarded due to their very short duration. As shown by Keeler (1972) through his work with organ pipes, attack and decay sections of notes are harmonically less stable. Thus, in order to analyse only the stable part of the notes, we use the sustained middle third of each note instance. This was located based on manual annotation of note onsets. When using large amount of data, the onset detection and note transcription could be performed automatically as in our recent studies (Köküer et al., 2014) (Köküer et al., 2014).

2.2 Modelling

An acoustic model is created for each player k and each note n, denoted as $\lambda_{k,n}$. The modelling in this paper is performed using a single multivariate Gaussian distribution with full covariance matrix. The parameters, the mean and the covariance matrix, of each model are estimated using the training data.

We consider the identification of players from a finite set of players based on a given piece of recording from the test data, containing one or more instances of a given note. The feature extraction step, as described in Section 2.1, provides a sequence of feature vectors. Considering that a given piece consists of R instances of a note, we have a set of feature vectors $O=\{O^{(i)}\}_{i=1}^{R}$. Each note instance is represented by a sequence of feature vectors $O^{(i)}=(\mathbf{o}_1^{(i)},\ldots,\mathbf{o}_{T_i}^{(i)})$, where T_i is the number of frames in the instance i and $\mathbf{o}_t^{(i)}$ is an N dimensional feature vector for frame t. The overall probability of the feature set O is calculated on model of each player k for the corresponding note n as

$$p(O|\lambda_{k,n}) = \prod_{i=1}^{R} p(O^{(i)}|\lambda_{k,n}) = \prod_{i=1}^{R}\prod_{t=1}^{T_i} p(\mathbf{o}_t^{(i)}|\lambda_{k,n}) \quad (1)$$

and the recognised player k^* is found as

$$k^* = \arg\max_k p(O|\lambda_{k,n}). \quad (2)$$

3. EXPERIMENTAL EVALUATION

3.1 Data description

The recordings chosen for analysis are part of a corpus of flute melodies, selected from commercially available sources and assembled under an AHRC Transforming Musicology project (Köküer et al., 2014). They feature the solo flute playing of Harry Bradley, Matt Molloy, Conal O'Grada, Séamus Tansey and Michael Tubridy who are prominent musicians in Irish traditional music. From the available solo unaccompanied recordings, four traditional tunes were chosen for each of the five players (note that the tunes are not the same across the players). The 20 tunes vary in length, from 17 to 41 seconds, and each tune contains between 147 and 311 events, including notes, ornaments and breaths.

3.2 Manual annotation

The original recordings, sampled at 44.1kHz with 16 bits and stereo, were converted by summing the channels to mono audio. Manual annotation of the recordings was performed by an experienced Irish flute player using Sonic Visualiser (Cannam et al., 2010), along with the Aubio vamp plugins Pitch Detector and Note Detector (Brossier, 2006). The annotation provided the start and end times of each event, the type of event (note, ornament, breath) and the F_0 fundamental frequency (Köküer et al., 2014).

3.3 Experimental setup

The audio signal was analysed using frames of 1024 samples with 256 samples shift between adjacent frames. The windowed frames were zero padded to 2048.

We employed 4-fold cross-validation to obtain more statistically reliable assessment of the performance. For each fold, three recordings from each player were combined to estimate the model for each player and note. The remaining one recording from each player was used for testing. This training and testing was repeated four times, taking a different subset as the test set each time. Overall test performance is aggregated over the four folds.

3.4 Harmonic analysis

We first present a visual assessment of the relationship between magnitudes of harmonics. For each instance of a note in each recording, the mean magnitude value of each of the four harmonics over the sustained middle third of the note instance signal frames, as described in Section 2.1, is calculated. Figure 1 depicts two dimensional scatter plots

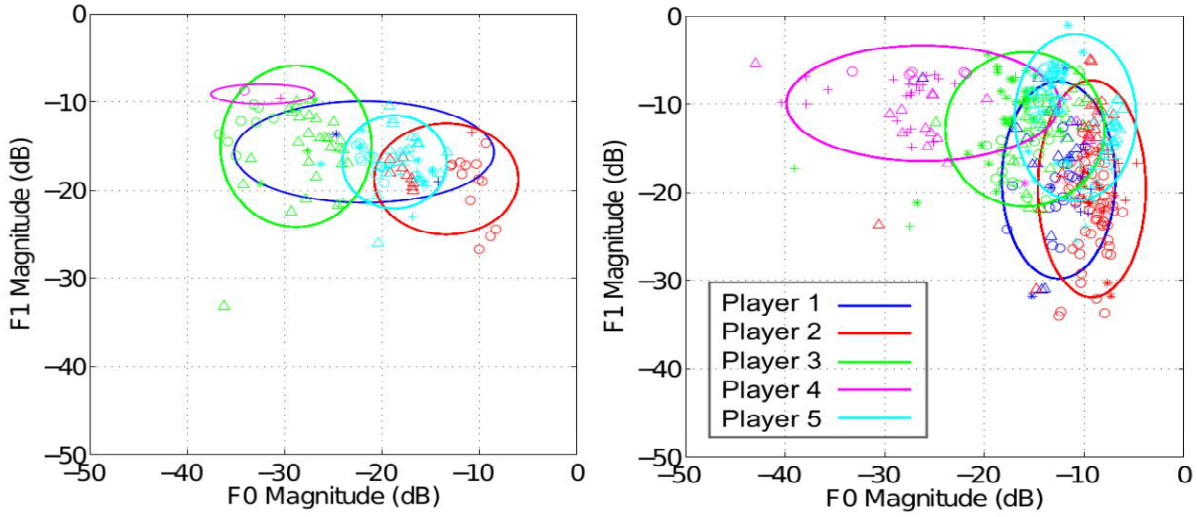


Figure 1: Scatter plots of the short-term magnitude values (in dB) at the harmonic peaks F_0 and F_1 for note E_4 (left) and note A_4 (right). Individual players are indicated by different colour and tunes by different shape markers. Clusters belonging to each of the five players are marked by a two standard deviation ellipse for visual reference.

of the mean magnitudes of one harmonic component (F_0) against other harmonic component (F_1) for all instances of note E_4 and A_4 from all recordings. In the figure, individual players are indicated by different colour markers and four tunes belonging to each player are indicated by different shape markers. For an easy visual reference, clusters belonging to each five players are marked by an ellipse showing two standard deviations with corresponding colour code. The clusters depicted in Figure 1 indicate that Player 3 (cyan) often exhibits similar levels at F_0 and F_1, whereas Player 2 (red) has a consistent F_0 with a variance in F_1. Player 4 (magenta) often exhibits F_1 magnitudes that are higher than F_0. These trends are evident across notes in the first octave (D_4 to B_4). We observed that in the second octave these trends are much less obvious with the exception of $F\#_5$.

3.5 Identification results

This section presents results of player identification. Table 1 shows the overall player identification accuracy when using data from individual notes. Results were obtained based on using single, two and three instances of notes. The accuracy was computed as the percentage of correctly identified instances out of the total number of instances over five players and over four folds. It can be seen that the recognition accuracy usually improves slightly as the number of used note instances increases. There are few exceptions to this, mainly in the case of using 3 note instances (i.e., the last column in the table). The decreased accuracy in these cases may be due to a note instance that is being modelled poorly by the correct player model affecting negatively the recognition outcome in more cases than it would have if basing the recognition on single note instances. It can be seen that the highest accuracy is achieved for note E_4 which is 57%, 63% and 68% respectively when using single, two and three note instances. The second best results with the accuracy of 51%, 54% and 55% are achieved for note A_4. On the other side, using data from note $F\#_4$ achieved the lowest accuracy, 19%, 22% and 26% which is at the border of or only little above the random guess.

Player Identification Accuracy (%)			
Note	Number of note instances		
	1	2	3
D_4	35.8	38.0	34.5
E_4	57.1	62.9	67.7
$F\#_4$	18.8	22.2	26.4
G_4	41.9	40.0	37.5
A_4	50.8	53.7	55.2
B_4	44.1	49.6	48.8
D_5	36.4	36.7	37.3
E_5	40.6	40.3	39.2
$F\#_5$	47.0	49.7	46.7

Table 1: Player identification accuracy (%) for each note when using data of one, two or three note instances.

Figure 2 shows the confusion matrices obtained when using the best performing notes E_4 and A_4. These results are obtained for single note instance and counts are accumulated across all 4 folds. In the confusion matrix, each row refers to the actual player and column to the predicted player. It can be seen that in the case of E_4, there are only few note instances for Player 1 and Player 4 which may have artificially positively affected the overall results for this note. In the case of A_4, the best identification performance is achieved for Player 3 and Player 4.

Table 2 shows identification accuracy across the 5 players when using all types of notes. It can be seen that Players 1, 3 and 4 are identified with considerably higher accuracy than Players 2 and 5.

E4		Predicted Player				
		1	2	3	4	5
Actual Player	1	**0**	1	0	0	2
	2	0	**4**	0	0	19
	3	0	0	**26**	0	14
	4	0	0	2	**0**	0
	5	0	12	1	0	**38**

A4		Predicted Player				
		1	2	3	4	5
Actual Player	1	**29**	13	7	0	0
	2	29	**43**	2	0	26
	3	9	3	**61**	10	12
	4	0	1	6	**34**	3
	5	0	33	30	0	**23**

Figure 2: Confusion matrices for identification of players, using single note instance, for notes E_4 and A_4. The rows refer to the actual player and the columns to the predicted player and the values represent counts.

Player Identification Accuracy (%)			
	Number of note instances		
	1	2	3
Player 1	51.1	57.9	58.5
Player 2	31.3	30.1	28.1
Player 3	49.1	53.4	55.4
Player 4	56.7	53.8	50.6
Player 5	32.6	35.3	36.3
Average	44.2	46.1	45.8

Table 2: Player identification accuracy (%) obtained for each player when using data from all notes and one, two or three note instances.

4. DISCUSSION AND FUTURE WORK

This paper presented our initial work on identification of professional players in Irish traditional flute music using short-term spectral magnitudes of harmonics as timbral descriptors and Gaussian classifier.

It was shown that the timbral variations between professional players are small but often show differences that are identifiable across a range of different notes. Players at this level often show F_1 harmonics at the same level or in excess of F_0. There are variations between harmonic levels on individual notes but at this level of playing timbre depends on the magnitudes of the higher harmonics in comparison to the fundamental (Fletcher, 1975; Pollard, 1996).

The player identification performance of the current system is rather low, which may be due to several factors as summarised below. The use of a single Gaussian distribution for modelling may not be sufficient. A more complex model, e.g., Gaussian mixture model, will be employed in our future work. The amount of training and testing data for some notes was very small. This is due to the recordings being in a range of keys which has a bearing on the occurrences of certain notes within the tunes. The audio we used here is from commercial recordings. This has an advantage of being available in the public domain, however, the disadvantage is that there is little specific information with regard to room acoustics, choice of microphone or equalisation added during the recording and mixing process. In addition, the players may have used different makes of flute instruments. In future work we will use a larger dataset that incorporates multiple repetitions of each tune played by the same player to minimise the effect of different instruments and recording techniques. We will also study an employment of compensation techniques, working either in the feature or model domain, to compensate for these effects.

Recreating the tests with control over the recording process and selection of corpus would result in a study showing a more realistic comparison between players and allow for deeper study into individual players' timbral treatment of specific parts of tunes as well as discovering whether consistency exists across a variety of different tune types.

Looking in detail at individual notes in line with Schouten (1968)'s acoustic parameters would also extend this research. In particular, analysis of attack phases and changes in spectral envelope during notes would be of interest.

5. ACKNOWLEDGEMENTS

This work was partly supported by the Arts and Humanities Research Council (AHRC) under the Transforming Musicology programme.

6. REFERENCES

Ali-MacLachlan, I., Köküer, M., Jančovič, P., Williams, I., & Athwal, C. (2013). Quantifying Timbral Variations in Traditional Irish Flute Playing. In *Proc. of the Third Int. Workshop on Folk Music Analysis (FMA)*, Amsterdam, Netherlands.

Breathnach, B. (1996). *Folk music and dances of Ireland*. London, UK: Ossian.

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

Burred, J. J., Robel, A., & Sikora, T. (2010). Dynamic spectral envelope modeling for timbre analysis of musical instru-

ment sounds. *IEEE Tran. on Audio, Speech, and Language Processing*, *18*(3), 663–674.

Cannam, C., Landone, C., & Sandler, M. (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proc. of the Int. Conf. on Multimedia*, (pp. 1467–1468). ACM.

Chudy, M. & Dixon, S. (2010). Towards Music Performer Recognition Using Timbre. In *Proc. of the 3rd Int. Conf. of Students of Systematic Musicology*, Cambridge, UK.

Coltman, J. W. (1968). Sounding mechanism of the flute and organ pipe. *The Journal of the Acoustical Society of America*, *44*(4), 983–992.

Erickson, R. (1975). *Sound Structure in Music*. University of California Press.

Fletcher, N. H. (1975). Acoustical correlates of flute performance technique. *The Journal of the Acoustical Society of America*, *57*(1), 233–237.

Hamilton, S. C. (1990). *The Irish Flute Player's Handbook*. Breac Publications, Eire.

Handel, S. (1995). Timbre Perception and Auditory Object Identification. In B. C. J. Moore (Ed.), *Hearing*. Academic Press.

Jensen, K. (1999). *Timbre models of musical sounds*. PhD thesis, Department of Computer Science, University of Copenhagen.

Keeler, J. (1972). The attack transients of some organ pipes. *IEEE Trans. on Audio and Electroacoustics*, *20*(5), 378–391.

Köküer, M., Ali-MacLachlan, I., Jančovič, P., & Athwal, C. (2014). Automated detection of single-note ornaments in Irish traditional flute playing. In *Int. Workshop on Folk Music Analysis (FMA)*, Istanbul, Turkey.

Köküer, M., Jančovič, P., Ali-MacLachlan, I., & Athwal, C. (2014). Automatic detection of single and multi-note ornaments in Irish traditional flute playing. In *Proc. of the 15th Int. Society for Music Information Retrieval Conf. (ISMIR)*, (pp. 15–20)., Taipei, Taiwan.

Köküer, M., Kearney, D., Ali-MacLachlan, I., Jančovič, P., & Athwal, C. (2014). Towards the creation of digital library content to study aspects of style in Irish traditional music. In *Proc. of the Int. Workshop on Digital Libraries for Musicology*, London, UK.

O'Neill, F. (1998). *O'Neill's music of Ireland*. Pacific, Missouri: Mel Bay Publications.

Pollard, H. (1996). Timbre and Loudness of Flute Notes. *Acoustics Australia*, *24*, 45–46.

Schouten, J. F. (1968). The perception of timbre. In *Reports of the 6th Int. Congress on Acoustics*, volume 76, (pp.Ì0).

Timoney, J., MacManus, L., Lysaght, T., & Schwarzbacher, A. (2004). Timbral attributes for objective quality assessment of the Irish tin whistle. In *Proc. of the 7th Int. Conf. on Digital Audio Effects (DAFX)*, volume 1, Perugia, Italy.

Widholm, G., Linortner, R., Kausel, W., & Bertsch, M. (2001). Silver, gold, platinum - and the sound of the flute. *Proc. of the Int. Symposium on Musical Acoustics*, *1*, 277–280.

Williams, S. (2010). *Irish Traditional Music*. Abingdon, Oxon: Routledge.

TOWARDS A NEW HEARING OF OF MUSIC FROM THE WORLD OF ORALITY: THE CASE OF SUB-SAHARAN AND CARIBBEAN-GUYANESE MUSICAL CULTURES

Apollinaire ANAKESA
CRILLASH EA 4095
Université des Antilles
Apollinaire.anakesa@free.fr

ABSTRACT

Music is a multifaceted, pluridimensional notion, with complex dimensions. It is governed by the written tradition and/or oral which defines it creation of rules, but also its practices, usage, symbols, meanings and stakes. It is a form of organization of social behavior and communications. It is a culture of life, but also social representating, sharing and exchange. Of its shares, references and many achievements result all kinds of speeches and descriptions, through which broadcast also varied musical and extra-musical knowledge. By this knowledge also define the characteristics identities. How to translate such a polysemic reality? To date, analyzes of scholarly musical thought westernized tools to study the prevailing various world music. These analytical methods are not always suitable to elucidate the true nature and reality of oraliture music. Attempting to meet this challenge requires that we dedicated to new analytical experiments. This is the task that I am working here, through the study of cases from the sub-Saharan and Caribbean Guyana music, which will serve me of illustrations.

1. INTRODUCTION

In most of the World cultures, music is a multifaceted, pluridimensional notion, with complex dimensions. Depending on place and context, oral or written tradition are defined its essence and the creation by specific meaning and uses. For instance, in the case of Western written music, the context of production and diffusion can be related to a historically defined place, especially related to the use of written documents which allow us to examine how music is made. The tools for its analysis are also well defined, if still under going modifications.

In musical systems grounded in oral tradition, as in Sub-Saharan Africa and the Caribbean, music is made through practice and a creation process which can be called "3D analysis". It is a combination of several layers of facts, acts, and interpretations, which can be seen from many perspectives, but still remain in strong interrelation. Music in these areas is never a single, simple event or fact. It is always a combination, but not necessarily a construction which could be described from its elements and their relations. The network of relations between music, the musicians, and the cultural environment gives birth to music as a total event. It includes sound, gestures, movement and instruments, with emotion organizing it as a whole, in building relations links, and in filtering, feelings, and memories through sensations. Actions and reactions as well as discourses are just a small part of the knowledge about music, in describing some of its characteristics.

The performanceof music is therefore often a complex of traditions and singularities, built on a socio-cultural system where the art of saying through music is a common and specific basis. Each act of creation, each performance, has its own dynamics, its own origin and goal, its own effect and meaning: day-to-day or sacred, tangible or imaginary realities. The musical "ideas" which emerge one proffered, not just to be heard, but also to be seen and verbalised. Meaning is only achieved through a process which is both collective and individual and which includes perception and culture, extratemporal dimensions and living movements.

To show the structural dimensions of this moving complex, such correlations and effects can not just be enumerated. One needs to cross the various layers at various angles, and this will be achieved by pluridisciplinary methods and tools. The position of the observer must also be included in the process.

The challenge that the analysis of African and Caribbean music offers to analytical tools developed for Western written music is a perfect occasion to renew presents general knowledge and methods. A series of case studies will offer a look into this new perspective.

From this point of view, it is proves necessary first to clarify the essence of these musical expresssions. I will start from the general move on to the specific, to reflect the global context and the representation of music as a phenomenon which encompasses a multifactor notion and complex dimensions.

The second step will allow us to understand the process of creation and the practice of concretisation of a work, a stratified process that I called stereotomic.

The third step will ask questions about the analysis we can do and with what means. And, lastly, for what result.

Given the magnitude of the task, this article will only refer to a partial and experimental conclusion. Further research, we hope, will provide more definitive answers.

2. MUSIC?

Today, music is a generic noun used in the singular, and in an almost universal way, to describe which reflete a reality both artistic and cultural. Complex, as music is indeed a universe of interdependencies based on sound frescoes that are submitted to the coherence of diverse realities and to the sum of multiple production experiments. These latter ones involve also a variety of prota-

gonists: individuals or groups of individuals, community, society, nation or civilization.
Music is consists of analogies of parameters that are multiple (technical and esthetical, physical and metaphysical, philosophical or ideological). Their functioning stays linked to the balance between the unity and the diversity of components that generate them inside a given culture and era, in specific contexts.
This set of factors intertwine and clarify one another to give meaning to the produced musical work. The act that guides this musical work then becomes, in essence, a political and social act in the wide meaning of these terms. Through this act, and through the underlying behaviors and practices, the individual takes part in social life, music then becoming for him a way of sharing and exchanging as well as of communicating and interrelating. Through music, soothing, to relieving, irritating, transcending, strengthening and other emotional acts (physical or intellectual), are then possible.
So, through music, man acquires as well his power of action and life as imagination, dream and vision. It is used as an act of individual or collective pleasure. It gives him a physical or spiritual comfort. It allows him to communicate with his kind as well as with other beings, including extrasensorial powers. It is also used as a way to transmit knowledge and values.

3. FROM THE NON-EXISTANT TO THE FINISHED WORK: TOWARDS STEREOTOMIC MUSICAL CREATION

"Man does little but invent" Paul valéry said (1944:87) about human creation. And in the field of arts, music is made through a savoir-faire and a creative process that I will call stereotomic invention. Stereotomic, as it involves, as already mentioned in the introduction, a stratification of data or layers of musical and cultural factors with plural elements. These elements, both of form or of substance, remain in a strong interrelation, and so allow multiple readings.
This is a process ruled by various musical gestures, fruit of the interconnexions and of the equally varied interactions, which, expressed in dynamic sounds and movements, are carried out in the time and space of the "Tuning of the World", in the words of R. Murray Schafer (1977).
Thus occurs a process of blossoming of the sounds acting inside man or surrounding him. Captured, perceived and organized through his artistic experience, our artistic creators explore the different possibilities of these sounds to give them a musical meaning. To do so, several approaches are possible and sometimes necessary. This is particularly true about the awareness allowing an analysis that gives meaning to the significant specific elements, as well as bringing to light the substantialty and the coherence of their entity. Such an analysis is necessary even for the creator of music. It elucidates the unities through which he conceives the appropriate style of language, and on which he founds the reality of his work. This language is ruled also by the way of thinking and the values that are the basis of the culture (individual and societal) of the artistic creator. The musical meaning that his creation will take on depends on these ways of thinking and these values. In a more general way, the specificity of the musical creations, as well as the choice of genres, styles, esthetics, and even philosophies and conceptions that one may have of a specific musical work also depend on them. The same is the case with the technical and expressive means that are used (language, gestures, graphics, signs, symbols, representations, behavior etc.) as well as the aim of the work produced.

4. TO ANALYZE? WHAT, HOW AND WHY?

The limited scope of this paper will certainly not allow me to exhaust the question on the series of questions raised both here and in the introduction: what are the relevant musical readings that can be adopted for the cultures that I use as an illustration, and for what analysis and to what end? What is the position to be adopted by the analyst, and for what observations?
However, preliminary observations are an indispensable precondition to open research pathways that, later, will lead to meaningful answers. These observations on the subject wath I started now, that should be developing in the future, will be based this time on examples from the Sub-Saharan and Caribbean-Guyanese musical culture.
As with all other musical expressions of the world, let's first remark that the cultures that rule them are protean.in their essence as well as in their achievements, these Sub-Saharan and Caribbean-Guyanese musical cultures have close links of interrelations and interdependances, that are stratified and complex. Thus they constitute a real cultural sound stereotomy.
Furthermore, the notion of music that these cultures express does not have –in the locally used languages- a single noun that would depict or translate only the exclusive meaning: the art of organizing technically and esthetically the sounds (cf. Anakesa 2007). However, a word or a group of words mostly express a vast idea and a multilayered concept, through interrelations and interdependdances combined or overlapping. In general, this concept essentially underlies the very close link between music and dance events, the Word (linked to the human language but also to the language of the spirits), and all sorts of other extramusical interconnections.
So the musical act takes on a communicative importance that is comparable to the one of spoken speech. As a corollary, the musician is supposed, not to play- in the first meaning of the term- his instrument but to "make it speak", to make it "dialogue" or "say things".
Here, music is the parallel of the spoken word and the dance. Music also serves as a true metalanguage modeled on the spoken language. The musical's codes and discourses have value of the spoken word, and the underlying meaning then becomes the musical verb to be exploited for all kinds of exchanges and communication. These latter are archeived through musical instruments

(voice included) that are not merely thought in their simple organological and acoustic nature. So, zoomorphic, anthropomorphic or the ordinary instruments are raised to the rank of cultural beings, equivalent to humans and to the spirits of ancestors or deities etc. Their disposition at the core of the group or on the stage is made in consideration of their "social" status, but also of their purely musical role inside the orchestra. The sound of each instrument, too, is a major factor in this real and symbolic consideration and organisation of musical achievement.
Here, the anchoring sound materials simultaneously form semantic elements. They also know an atypical structure. Combinations of a double expression coming from spoken language come from it. This typical and coded expression respects a structure with an appropriate grammar. Musically, arrangements are made, which are as singular for the style, or the aesthetics. All these musical expressions are conceived as being the human experience that expresses the order of inner life. Therefore, music is here one of the major factors that contribute to the regulation of this life and allow man to have a connexion with the environment, be it cosmic, natural or human.
All this justifies why, among other things, Sub-Saharan and Caribbean-Guyanese musical productions are at the same time the echo of social organization and the set of underlying roles.

5. AN EMPIRICAL READING OF THE MUSICAL REALITY OF THE SUB-SAHARAN AND CARIBBEAN-GUYANESE ORALITURE

Before proceeding with my analysis, I would like to say that the use I make of the concept of oraliture is linked to the meaning given by Patrick Chamoiseau (1992:425), who defines it as everything that, in the order of discourse, but also of culture, is linked with oral and not written litterature. Is it necessary to say that, nowadays, oral and written combine and are of equal value, as much as for what they mean individually as for the effectiveness of their respective systems of organization. This system allows for significant communications, able to translate all kinds of realities and representations.
Let's go back to the study of cases that concern us, and let's consider to begin with the *Kasékò* of the Guyanese Creoles. Among them, this word, that litterally means "to break the body", is a word that is linked with a reality of multiple interdependencies. Indeed, this one word refers at the same time to songs, entertainment music, dance, musical instruments and the night or the time of the performance when they are practiced. By extrapolation or by analogy, this word refers as well to values, a system of a way of thinking and of life, a way to share and communicate, but also to a system of alliances that echoes a peculiar organization of the Guyanese Creole society. *Kasèkò* thus consists of a stratification of musical and extramusical interdependent parameters. These parameters operate under several levels of relationships, either hierarchic or egalitarian, but in all complementary. The production space of a *kasékò* musical performance, called *swaré kasékò* or *bal kasékò*, has four major music scenes:

The space of singers, predominantly women. It is ruled by *Larèn* (the queen, main female singer and soloist). Her special status allows her to play the role of conductor, and her stick is the *chacha* rattle that she uses wisely and in timely fashion.
The space of dancers (male and female). It is ruled by a commandeure, a kind of foreman who, through the circumspect strikes of his *tapèt* (two wooden pallets, with a square shape) announces the start of the choreographic intervention, and leads the steps while directing the movements of the dancers (male and female). He also announces the change of choreographic figures, as well as of the couplets or the tune to the female singers, and even of the tempo to the whole orchestra, or the end of the party, etc.
The space of musicians, traditionally male, even if, women participate more and more. Here, the dòkò, the drum master, is king. He plays the solo drum. He is remarkable for his playing and his playful (virtuoso) improvisations.
The space of the audience, and its extension, *the public restaurant space*. Here, meals, alcoholic and non-alcoholic drinks feed the body, allowing individuals to stand for the duration of a *swaré kasékò*, which start at 9 pm and end the next morning around 7 am.
These spaces are not partitioned. They are interrelated, and the individuals who compose them circulate inside, going freely from one to the other, while respecting the rules of this musical representation. So, they can become in turn, audience members, musicians or singers.
Nowadays, this music is more and more musical show and concert during various festivities (festivals, wedding ceremonies or other occasions of entertainment, public or private). The configuration described above is then adapted to the context and the circumstances.
Musically, the singing has a responsorial structure: the choir answers with the refrain the soloist who sings the verses. The singing is generally monophonic, but may, because of the diversity of the vocal registers of the members composed of Youngs and adults, cover a heterophonic tone.
Among instrumentalists, the currently most common style is the one of the *kasékò* with three drums: the *tanbou koupé* (solo drum played by a master, the *dòkò*) is in the middle. To the left is the *tanbou plonbé*, an accompanying drum with a low tone, equally tanbou bass whose role is to strengthen the harmonies and the tone of the ensemble. On the right of the solo drum, the second assistant plays the drum with a medium tone, the *tanbou foulé*. To the left of this drum is the idiophonist *bwatiè*, the *tibwa* player. This idiophon is made of two hardwood sticks and a stool-shaped wood drum, which the musician strikes. It gives the tempo, and zest to the instrumental ensemble.
Tanbou plonbé, tanbou foulé and *tibwa* play short, and repetitive motives with a quite dynamic *ostinato*. Through their respective instrumental playing, they keep a strict relationship of complementarity. To maintain the harmony of the ensemble, each one carefully listens to the neighbouring musician and does not lose the harmony of the group. They look as well at the gestures and facial

expressions of the neighbour, to which, through a complementary instrumental playing, they must answer by meaningful and challenging movements. Listening, hearing and seeing are thus capital acts to be observed. They allow for complicity in the individual instrumental playing and the partners. They enliven the imagination of all, and give sense to the values, the symbols and the representations underlying the *kasékò* musical show. Special attention is given to the instrumental or vocal tone, symbolically able to represent the voice of a human being, an animal, or an ancestor or a deity when one is in an initia-tic or ritual context.

Figure 1. Guyanese *tanbou kasékò* and *tibwa*

Figure 2. Guyanese *tanbou yangwé* (male & female)

Figure 3. Martiniquese *tanbou bèlè* ©Simone Vaity

Figure 4. Guadeloupean *tanbou ka* (*gwoka*)

Figure 5. *Lewoz* performance with *tanbou ka* (Guadeloupe)

Figure 6. *Kasékò* performance (French Guyana).

There is a hierarchy of position and role in the musical organization of the *kasékò*. The *dòkò* is at the centre of the system, a master musician with multiple and varied skills (auxiliaries musical and extra-musical knowledge).
As for the *kasékò* dance, women and men are engaged in a constant game of seduction. The steps and the basic dance movements function lascivously, with a repetition

with slight variations for the women. The men show their prowess, especially through the *nika*. In this style, all kinds of acrobatics, including the imitation of the postures of animals, add to the beauty of the choreography and embody the first meaning of this music: "to break the body".
Two major technics govern the sound organization mode of the *kasékò*: the repetition by all kinds of cyclical *ostinati*, and the improvisation of the soloists (singer and instrumentalist), but also the prowess of the dancers. Inside this organization, in the systematic analysis areas, appear melodico-rythmical practices built in short cells. Syncopations and back beats alternate and/or overlap inside one or at least two meters. They contribute to polyrythm. Despite their repetition, the slight variations inside the rythmical cycles increase the vitality of this repetitivity, and avoid the impression of weariness of a less dynamic repetition.
This music, born in the universe of the slave plantations, is steeped in history. It has its roots in Africa at the origin of the ancestors who gave birth to it.
This results in a relation linked or to reflecting the social values of respect and consideration of others, of the care with which each dance movement, instrumental play, greeting, courtesy and respect of the rules govern the function of each one of the underlying actions of this musical activity as a part of the basic values. They belong to music and society, culture and identity.
Many things remain to be said about the subject of *kasékò*, and many as well about the following examples. I will treat them in new ways, but whichare complementary, to follow my reflection about the questions raised in the introduction, as the limits of this article do not allow me long developments. So I will shorten my point, by restricting myself to the necessary minimum for the relatively better known cultures (like the Martinique *bèlè* or the Guadeloupe *gwoka*). I will thus be able to give a little more detail about the *bushinengé* culture, or that of the Maroons of Guyana.

6. FROM MUSICAL REALITY TO SUB-SAHARAN AND CARIBBEAN-GUYANESE CULTURE PATTERNS: MUSIC AS AN ART AND CULTURE OF CONCEPTION, COMMUNICATION AND LIVED SOUNDS EXPERIENCE

In the Sub-Saharan and Caribbean-Guyanese universe, music is in no way a word restricting the meaning of art to only the technical and esthetical factors, like the dominant Western thought which, in this regard, is tending to become universal.
In Sub-Saharan Africa and in Caribbean-Guyana, music is everywhere woven into the social web of daily life. The production that comes from it thus forms a system of nested social relations and related to nature and sacred. At the core of their practices, and beyond the technical principles of musical creation, are interelations of distinct social relations (gender, generational, political, festive, ceremonial, magical-religious or festive).
In Sub-Saharan Africa in particular, the diversity of the languages or of the sound systems is a reflection of its numerous traditional societies, hierarchical or not. The major part of their musical processes resemble rituals. Here, sound combinations contain spetific prototypes or features often linked with scales, rhythms and dances, to which refers the produced genre. That is, every musical system has an organization that depends on an inner logic very often conditioned by external factors.
Technically, the structure of the parts often rests on little simple melodic and rhythmic units. Some are repetitive, others include variations, notably through improvisation. During the development of the musical discourse, generally based on a pivot, the variation of these basic elements may be of intervallic, rhythmic or accentual order that have an effect on the periodicity of length and metric. The changes also concern tones, one of the essential parameters of Sub-Saharan traditional musical systems.
In the case of vocal, instrumental or mixed group, all of these elements contain synchronisms and criss-cross producing complex combinations. The musical data are then often linked with the socio-religious circumstances to which they are integrated. The cultural data are often linked with symbolism. They remain meaningful and also bear out strictly technical musical criteria.
This also explains the fact that the names of the musical genres derive from denominations of rituals, ceremonies, instruments or other contexts and social interactions, as types of communication that are transmitted musically in association with some event, symbol or ideology. For example, puberty rites of the Aka of Ghana have the same name as the musical genre that is associated with them: *dradwom* (puberty songs).
To name a ritual musical genre, for instance horn music, one can add a word before of after this name to determine the nature of the ritual and the kind of horns used (animal tusk or wood).
So, be it voice, instruments or their combination, the African musical sound systems include sounds with a determined pitch as well as those with an undetermined pitch, often with a functional character. These sounds may sometimes be symbolically linked with the voice and the rank of individuals or of the spiritual beings. So we may encounter instruments producing male or female sounds, sounds associated with the voice of ancestors and deities, the voice of father or a mother etc.
Although unwritten, the musical expressions of Black Africa include very sophisticated sound textures as well as rather complex structures. These textures are generally

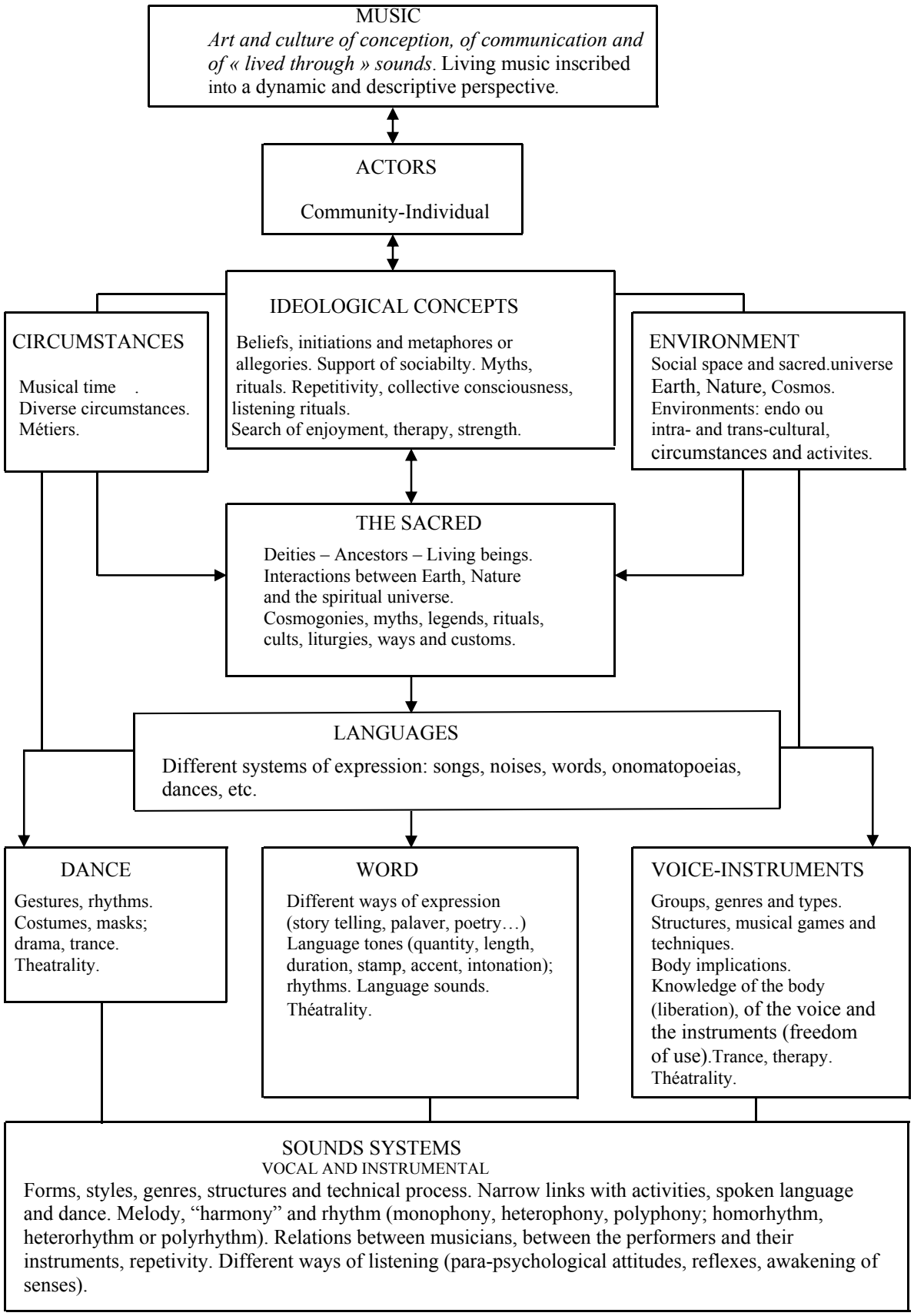
rich in tones, thanks, among other things, to the accessories or surrogate that performers have. These accessories are generally attached to the forearms, wrists or feet or attached to the main instrument.

In Sub-Saharan Africa, as in the Caribbean-Guyana, the musical organization is conceived as a chain with multiple links (musicians, singers, dancers, audience), interlinked and having a keystone, which can be a leader female singer as among the Creoles, or a master drummer among the Bushinengé and the Sub-Saharans. Through their gestures and their performance, they produce a chain of speech the value of the meaning of which is similar to that of the spoken language. In this context, some instruments (anthropomorphic or zoomorphic or wearing some symbolic representation on their body) produce sounds that are assimilated either to the human voice, or to that of an ancestor or a deity represented by the instrument in question, the ancestor or deity is also able to possess a person in trance, through whom they will express their voice.

Music is at the same time a produced sound, conceived and constructed by an individual or a community, but also a received sound, perceived and interpreted by them through significant and varied representations and symbolisations. Musical instruments in particular are conceived and used following the logic which they refer. A traditional group is constituted of instruments that are generally arranged precisely on the stage. This disposition reflects as well technical factors (according in specific sounds and languages how music is made, but taking ito account as well the overall sound) and also cultural factors (social status, symbolic representation, etc).

The musical gesture serves as a bridge. It generally helps to produce coded languages. So we can find an ordinary drum or a sacred drum. Their sound may be a plain sound expression that delights the ear and touches the soul. It may symbolize the voice of an authority, an ancestor or a deity.

ABOUT AFRICAN MUSICAL THOUGHT
CONCEPT OF « MUSIC » IN BLACK AFRICA: SUMMARY TABLE OF THE INTERACTIONS OF MUSICAL AND EXTRA-MUSICAL FACTORS



SOME BASIC OF SUB-SAHARAN MUSICAL CONCEPT
(Plurality of society- plurality cultures)

I. MUSIC:
→ *Complex phenomenon with multiple factors and strata, concrete and conceptual, of human existence.*
→ *Art and Culture of* conception, communication *and* lived *sounds.*

II. CLOSE LINK
MUSIC - DANCE - SONG:
→ *Music = dance.*
→ *Music = word* (sung words, musical instrumental "verb" and "speech").

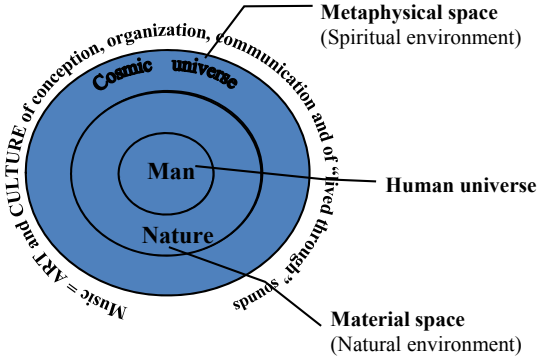
III. REPPORT
BODY – SOUND – VERB (SPEECH):
On physical and metaphysical sense

IV. MEANING OF MUSICAL INSTRUMENT:
→ *Culturel Being.*
→ The luthier *"give voice" to the instrument* and the musician *"make her speak".*

V. REPPORT
MUSIC – NATURE - COSMOS :
Sacralized musical universe
→ para-ritual music (*entertainment, trades and ceremonial*);
→ Sacred Music (*ritualized*).

ABOUT AFRICAN MUSICAL THOUGHT

1. Use of sound:
Broad concept involving fixed and indeterminate sounds, *feature the acoustic properties of sounds exploited; search and selection of hybrid sounds and enriched stamps; based on a musical evocation of a destination , ritual or not.*



2. Relation to the sound:
Direct impregnation
(sound received, sculpted and musically organized by man to express emotions, feelings and physical and metaphysical lives).

2. Sound production:
With the help of musical instruments, with direct reference to Nature and the metaphysical universe.

Two principles related to intrinsic values are the basis of

8 (a & b).

and self-assertion, ability and sound is rarely of no consequence. A musical performance is, indeed, a genuine area of creation and of free-

9.

10.

Figures 7-10: Sub-Saharan and Caribbean-Guyanese universe;

the production of a musical

Figures 7-10: Sub-Saharan and Caribbean-Guyanese universe; sanza (Zandé anthropomorphic *likembé*); *Mambila* drum percussion (zoomorphic idiophone, Cameroon); Luba performers (mutwashi dance, RD Congo). ©Anakesa for pictures 7 & 10; photos 8 & 9:
http://sanzaskyneblogsbe/archive/2011/07/11/bruneaf-2011-sanza.html

dom of expression that one enjoys in the respect of values and rules, for a social cohesion, a savoir-vivre in good harmony: to show the community and oneself to advantage. The value of the group generally takes precedence over that of the individual who gains thanks to the qualities of its contributions to that group.

The instrumental playing is materialized through the musical verb. Here one "makes the instrument talk" in the low, middle and high tessitura. To achieve this, one caresses or hits it, among other things. The Guyanese and Guadeloupean Creoles also do *plonbé, foulé* and *koupé* to produce the low sounds, obtained mostly thanks to the movement of the hand going from the top to the low part of the skin of the drum. The low and medium sounds are used as the base and support and accompaniment elements for the high tones of the soloist. Those high tones allowing him to distinguish himself, to put himself in highlight as a master.

In this space of creation, tensions and all kinds of emotions are released.

In Martinique, it is particularly in *bèlè* that music, dance and story-telling mix to transmit a certain conception of the world, an art and a culture of life, but, also, of communication by the art of sounds. Today, this art is realized through specific contexts, animated by songs, drum and some idiophones (*tibwa, chacha, labas, slyak, tryang,* etc), during the *swaré bèlè*. The context of performance or of production allows here gathering that is symptomatic of a way to be together, but also values of solidarity, of sharing and even of cultural resistance. *Bèlè* deserves its demonination that specially means a beautiful place, a good moment lived under the auspices of a nice air.

In Guadeloupe, it is the iconic *gwoka* that, through music, dance and song, represents history and identity but also playfulness, a certain way of being, of seeing the world, communicating and living the daily life. *Léwoz, menndé, kaladja, graj wulé (woulé)* and *tumblak (toumblack)* are the main rhythms and styles of *gwoka*.

Among the Bushinengé of Guyana, whose ancestors escaped from Surinam in the time of slavery, two major ceremonies are celebrated that results in the artistic event of great importance. They call them *pee*. They are funerary rituals, *booko dei* and *puu baka* (*pou baca*), entrance of mourning and lifting of mourning. Here, music and dance, that rule these events, result from the great dances organized before by the slaves of the plantations of Surinam from the XVIIth century. They are now structured in series of dance and story-telling.

The *mato* by which these ceremonies begin, is a session of storytelling featuring expert traditional storytellers. They compete amid a crowd gathered under the mortuary shelter, or at home, by the light of a lantern or a wood fire. During the ceremonies, historical, anecdotal, mythical tales and stories about ordinary social facts, often linked to the dead person, alternate or finish with the songs and the interventions of the drummers associed with the crowd. Singers, instrumentalists and audience react to the words of the storyteller, asking him questions, approving him or responding to all kinds of signals, including clapping hands, shouting, emiting spontaneous cries, doing some dance steps punctuated at times by gunshots in honour of the dead person. Then the music comes to a climax, before the drums become temporarily silent and give way to other story tellings that lead to other songs and dances.

The *susa*, successor to *mato*, is an acrobatic dance. Men dancers compete. Each one trying to perform the right movements and the acrobatic gestures that one must perform with dexterity and at the right moment. The rivalry takes place in a circular performance space, in front of the drummers. This circular stage is constituted by the musicians, the female singers and the crowd of spectators standing around them, among which the female singers who stand in front of spectators. The loser or the one "killed", (they say "*kii*"), leaves his place for the next competitor. It is through this dance that in ancient times the fighters exercised to defend the community from the slave masters who pursued them during their escape.

After the *susa* comes the *songé* (*son-ngué*), also named *agankoï*. It is a mixed dance where, on the stage, a group of women alternates with a group of men, with *kaway* cowbells tied to the ankles. Occasionally, a female dancer and a male dancer, alone or together, exhibit their prowess in front of the drummer with whom they interact through gestures and coded sounds. The lead singer stands out through melodico-rhythmic flights that generally end with a spellbinding *vibrato*. The soloist drummer launches into improvisations with coded rhythms of his musical speech. The musical meaning that comes from it has the role of spoken words, intelligible for initiates. As with the dancers, they compete in skillfulness and acrobatic demonstrations. For instance, with the help of their calf, they can throw up into the air a stool that they will catch between their legs before it reaches the ground. It is also at this time that the main drummer, always with coded formulas, may honour a distinguished and gifted dancer. The person concerned, in recognition, raises a hand over his head and shouts his recognition. Then he lies in front of the drums. He then stands up to make a demonstration of his style in front of the audience as a thank you. And if this demonstration is still better appreciate by the audience, the audience bursts into acclamations and exaltation.

The *songe* is preceded or followed by awawa. It is a session of songs originally dedicated to women, but in which men also participate. Through these songs, without accompaniment, are expressed all kinds of feelings and emotions that rule the daily life of the male-female relationships.

Here the song leaders succeed on music scene. Anyone can go on the frontline to sing a criticism, an insult or proclaim his rebellion or any other humoristic remark. Most often a man or a woman is accused of incompetence in performing household chores. Everybody has the right to speak, to answer, or to replicate other things to his rival, no connotation of animosity or revenge, everything happening in a friendly atmosphere filled with joviality.

The *awasa*, which closes this series of dances and storytelling, is a very dynamic finale. It consists of the demonstration and the virtuoso performance of stylized dance

steps – enriched by the rustling sounds of *kaway* cowbells tied to the ankles –, of the rhythms of the solo ,drum and the flights of a soloist male or female singer. This creates a special communication between the various protagonists, mixing musical codes and social codes, picturesque gestures and coded sounds, imitation of and reference to Nature, evocation or representation of the sacred.

This dance exalts gestual beauty and dexterity of movement of the male and female dancers, as well as the balance and the strength of attachment of their feet on the floor. Thus among other things, by pointing the toes toward the ground as to « plowing » dust, these dancers become light as cats, twirling like fighting cocks. The traditional *pangui* loincloths decorated with various embroidered geometric patterns that evoke the *tembé* art, gird the hips of the female singers, while men are dressed in kalimbé (traditional loincloths worn as a shorts), thus adding more beauty to the show.

All these musical expressions, dances and ceremonial story telling are accompanied by percussions instruments (three *doon* drums, one *kwakwa* idiophon, a *chacha* rattle and *kaway* cowbells). Musical instruments, certainly, but they evoke at the same time other beings and other relations. Their sound is in fact a « voice » and their melody and rhythm, speech, a stylized language rich in meanings.

So one may distinguish the *gaan doon*, great drum or master drum, that plays the solo part, establishing a dialogue with the rest of the musicians. The *pikin doon*, little drum, marks the rhythm and the master drum with its ostinatos which define at the same time the genre or style of the music played and the identity of their authors. The *tun*, the « watching drum », marks the beat and the tempo with regularly kept hits from the beginning to the end of the number, like a metronome. This latter drum may also play even more complex rhythms in some forms of ritual ceremonies like *sebikede, pudja* and *atompa*.

Played alone, the *gaan doon* is also used as a ritual drum with a specific language: the *apinti tongo*, the language of the ancestors' spirits and those of the deities. In this context, this big drum changes its name and then becomes *apinti doon*, *apinti* drum, sacred drum. Every ceremony begins with its rhythms and sounds, as a prayer or greetings addressed to ancestors and/or to deities, so as to obtain their blessing to participants.

The kwakwa idiophone is a long board on which several musicians six or more), using two sticks, play at the same time, each one a different rhythm. They are the power unit of the polyrhythm that will result.

The *kaway* cowbells are worn by the female and male dancers around their ankles. Underlying the steps of their dance steps, the sounds made by these bells, as well as the gestural elegance of the arms of the dancers, add to the sound and choregraphical richness of the ensemble and to the aesthetics of hybridation very popular among the Sub-Saharan and Caribbean-Guyanese universe.

The practice of the *chacha bushinengé* rattle is generally exclusive of the Obiaman or Sabiman: holders of knowledge who officiate in ritual practices.

Figure 21. *Pikin doon tun gaan doon*

Figure 32. *Agida* sacred drum

Figure 43. *Kaway* Cowbells

Figure 54. *Kuakua*

Figure 65. Bushinengé Youngs drumers

Figure 76. *Awasa* dance performace. ©Anakesa for all photos

7. CULTURAL ETHOS AND MUSICAL TASTE IN THE SUB-SAHARAN AND CARIBBEAN-GUYANESE UNIVERSE

The Sub-Saharan and Caribbean-Guyanese musical reality is governed by an ethos that becomes an organizing principle of the social practices and behaviors, through the art of sounds, thanks to which various interdepencies are articulated. This articulation is also done by means of interactions (on the social area), signifiçcations, symbolisations and emotional (on the ideas, feelings and desires area). The underlying cultural aspect is thus produced through adequate practices. Three major dimensions underlie the interdepencies and the interactions that come from it: “awareness of what is possible, conception of the normal and feeling of the sensitive, of the emotional”, to quote Remy/Voyé/Servais (1980:279).

So this music expresses a character, a way of being, a disposition of the mind, which are the basis, among other things, a way of being and of a certain conception of the world that rules them. It is as well the echo of a social configuration, a reference to belonging, an attachment, an identity.

During festivals or ceremonies, alone or collectively, music becomes the strong cement of the community. Through it an atmosphere of great conviviality and social cohesion is created. It constitutes the central moment and place for meetings that will initiate many relations and alliances.
At the same time, the opportunities for musical performances, such as the *swaré kasékò, bèlè* or *gwoka*, for instance, as well as being the contexts and the places of the musical practices, are real referents that evoke essential and existential factors such as identity, a way of thinking and of living. Beyond the pleasure that may come from the musical entertainment, the practice itself of music and its relative, dancing, prove to be a school for life and knowledge, and edification of the being through musical practice.
In the Sub-Saharan and Caribbean-Guyanese cultural world, music is thus a fundamental notion of social thinking. Mostly during the parties and *ad hoc* performances, among other things, music echoes the family and community organization. It is used at the same time as a place and a means of meeting and of political representation, in the wide sense of the word, under the authority of a physical or symbolic leader (a master instrumentalist or singer, an ancestor, a deity). From that comes a similar organization in terms of respect of the rules and of the social values and of all kinds of ability. It is the same with the learning of music and of instrument making, which generally take place in the family, to which are associated many social practices, the knowledge of nature, the exercise of the sacred.

8. WHAT CAN US ANALYZE IN THIS KIND OF MUSICAL UNIVERSE?

By proven principles and methods, analysis allows the decomposition of an organic whole into its parts, while making explicit the meaning. Thus can be distinguished, depending on the nature of the musical work, we can distinguish its different components mostly its tones and formantes, its rhythm and tempo, its themes and melodies, its modes and harmony as well as its forms ; the instrumentarium, the agogic elements, etc.
The means to achieve this are many and varied. For a precise purpose, it is thus possible to make use of musical writing and other schemas, sonagrams, photographs, videos or audio excerpts; new technologies and multimedia (with all kinds of software that allow a refined analysis of rhythmic, melodic, harmonic and tone components as well as other sound riches that were, until not long ago, still inexplorable). In this approach, one should not forget that, for a musical system of oraliture as the Sub-Saharan and Caribbean-Guyanese one, the factors of sound, the notions of listening and seeing are primordial. They concern both musical as well as extra-musical elements.
Ainsi les moyens d'analyse, à adopter, contribueront-ils à l'élaboration des modélisations qui répondent au besoin non pas d'une analyse comptable et mécanique, mais de celle capable d'une mise au jour efficace d'isolats significatifs, et porteuse d'une véritable explicitation des faits ou de la réalité musicale observés.
Thus, the means adopted shall not be used to perform a mechanical and accounting analysis. The methods to be adopted should instead be used to effectively uncover significant isolates. Consequently, this analysis must include an explanation of the true facts or musical reality observed.
For the music coming from the Sub-Saharan and Caribbean-Guyanese system, that constitute my examples, the units or the technical components that constitute them will only have real signification and analytical pertinence if they are understood individually, but mostly in a global way, to give a meaning to the musical realization concerned under analysis. To do this, we must take into account the previously mentioned interdependence, through associations of musical factors and of extra-musical or cultural factors that are linked to them and give them meaning. It may be a language, a myth, a dance, a dress, an image, a symbol, a significant refrent or any element of representation. We can then report on these elements, among other things, by processes linked with hearing (sound extracts) and with sight (a transcription, a chart, a graphic, a diagram, a plan, a photography or a video), depending on the necessity and the relevance of the analysis.
This approach is justified by the fact that, in their process of musical creation of oral tradition, the creators working inside this system do not think about the different parameters as technical isolates to associate or accumulate. They do not search, for instance, to work at first only on a scale, in some tonality, to find first in what meter will be played such kinds of melodico-rhythmic cells or what intervals need to be worked firstly and what transposing instrument should be used to create the desired music work. In terms of consonance notably, for them, the hybrid sounds and all the richness of formantes they represent, and also the signification linked to them, are harmonious and pure. Thus, for their musical realisation, they think more in terms of globality of the autonomous or isolated factors that have to operate in narrow interrelation and interconnexion. It is moreover the same for the learning that happens through processes of globalization more than by a strict individuation based on small details ; if necessary, groups of details are often formed to reach the famous globality, while isolated factors being aware of the importance that must have every detail present in the system must have.
What matters more is the functioning of solidarity and interdependence between these isolates which each advance in a relation of reciprocical dependence. Thus it is a realization of independence inside interdependence are created. Independence and interdependence are also made in a logical consistency, of the technical, esthetical, ideological and meaningful elements. The intrinsic value of each of these elements does not constitute only the representative value, but is also a changing reality within the whole. It is also true to the specific characteristics that make their distinct reality.

Thus, musical reality does not have, in there cases, a unique organ as a substratum, as it remains diffuse in the social environment, the natural environment and in the metaphysical universe. Each of these universes give it specific characteristics that constitute at the same time a distinct reality. This, is close to the theory of Durkheim (1893:46) when he speaks of "economic and social reality".
What can we conclude from all this of informations?

9. TOWARDS A PRELIMINARY CONCLUSION OF A NEW EMPIRIC READING OF THE MUSIC OF ORALITURE

As we have just seen, the music of the Sub-Saharan and Caribbean-Guyanese universes is at the same time a complex entity of interpenetrations of multiple links both internal and external, individual or communautary, physical and metaphysical that are inseparable.
As is the case in many cultures around the world, the music of this universe is at the same time an artistic art and a cultural and social art. It is as also a metaphysical phenomenon. Through this, the self and the other, the individual identity and the collective identity are defined. By this intermediary, exterior and more tangible relations are also determined: temporal and timeless, individual and collective, intracommunautary and intercommunautary, endo-factional and exo-factional. The result is what I call "musical endo- and exo-physyognomy" a face for oneself and a face to be shown to the other. These exterior relations have specific sound identities. These sound identities include choreographic characteristics. They present at the same time analogies linked to the body and to the status of the musical instruments. The tone of those sounds, a synomym of voice, evokes all kinds of represetations and symbols. Thus the acts of musical performance that result are distinct and distinctive, in relation with the activity, notably initiatic.
Indeed, in the Sub-Saharan and Caribbean-Guyanese universes, construction and reproduction of social relations are not only found in the structures that underpin them. They are as well closely associated with musical practices, ordinary or ritual, because, here, everything is thought to be sacred. Thus the music promotes, among other things, reproducing social categories and the underlying relations, as well as their values (kinship group, community links in particular). Thus music here acquires a peculiar resonance through very dynamic technical and aesthetic elements that promote the social integration of the individuals and a conception of the world through specific means and vision.
In this context, the musical riches do not reveal as autonomous systems which can understand the nature only than from specifically musical parameters. Indeed, their systems are a set of stakes and of games that are differentiated and contrasting. Trying to elucidate them requires that one observes, defines and interprets relations and interrelations as well as interactions that rule all aspects not only of the music (sound objects, musical instruments and their acoustics, technics of playing, structuration of the sound material, repertories, genres, styles) but also of the extramusical : the actors create them, inside their sociocultural and even natural surroundings, without forgetting circumstances, representations, associated contexts and other factors.
It remains to emphasize a postulate. The language of the analyst is certainly not that of the musician. How to name the same facts and things related to music, remain different depending on whether one is a musicologist, a composer or a performer. Reading each of them made on these facts and these things are also different. In this regard, the organological work of François Picard is significant with a meaningful classification of the instruments. It reveals the different readings that all these actors have on a same object.
We saw that, in the Sub-Saharan and Caribbean-Guyanese musical systems, a close interdependence exists between musical performance and some aspects of social organization. Some musical genres and forms are also serve as tools of expression of multiple realities games, with temporal and intemporal referents. One can bee clearly that in the *kasékò, bélé* and *gwoka* gatherings. Some words serve to illustrate this junction between the musical and the social through which reconciliations and exchanges one made, but also, in the heterogeneity, sharing, alternance, and even oppositions, complementarity and harmony of the individuals.
Here, as we have seen, making music is a major act of establishing and connecting of alliances. The musical performances then become regulators of social issues among men, but also generators of relations between human beings and the worlds of spirits and of Nature from which relations with the natural environment, the cosmos are created.
In this system, the music is perceived globally, in connection with a structuration of the dependencies and interdependencies relations.
However, the analysis proceeds by isolating the elements of an organic whole to observe and see the nature of those elements, identify concepts, relations, values, practices: define its effects and even the action and the causes.
For a good analysis of the music expressions of the Sub-Saharan and Caribbean-Guyanese oraliture systems that concern my research, I aim an articulation among its general principles, that allow a better decomposition of this organic world into its elements, and its complement, its synthesis in some way, that reconstitutes everything from these given elements to give meaning to it, to translate reality and therefore, elucidate it. I will be the aim of my future and analyses, furthering this direction of research.

10. REFERENCES

Anakesa Kululuka, A. (2007) *L'Afrique subsaharienne dans la musique savante occidentale au XX^e siècle*, Paris, Connaissan-ces et Savoirs.

Arom, S. (1976) *Le langage tambouriné des Banda-Linda*, Paris, Selaf.

Bariaux, D. & Demolin, D. (1995) « Naissance de la voix d'un tambour à fente chez les Mangbetu. Du geste de l'artisan à celui du musicien et du danseur », *Cahiers de Musiques Traditionnelles*, n° 8, *Terrains*, Georg éditeur, Genève, Ateliers d'Ethnomusicolo-gie/aimp, 105-113.

Bebey, F. (1969) *Musique d'Afrique noire*, préface d'A. Martel, Paris, Horizons de France.

Belinga, E. (1965), *Littérature et musique populaire en Afrique noire*, Paris, Cujas.

Blerald-Ndagano, M. (1996) *Musiques et danses créoles au tambour de la Guyane française*, Cayenne, Presse Universitaires Créoles/GEREC, IBIS rouge Editions

Bocage, E. (1987) *Langues et cultures guyanaises*, Cayenne.

Calame-Griaule G. & Calame B. (1986) *Introduction à l'étude de la musique africaine, La Revue Musicale, op. cit.,* p. 14.

Chamoiseau, P. (1992) *Texaco*, Paris, Gallimard, 1992, p. 425 **et suiv**.

Doucet, V. (1989) *Musiques et rites afro-américains*. *La marimba éclot dans les astres*, Paris, L'Harmattan.

Durkheim, E. (2008) *De la division du travail social*, Paris, Les Presses universitaires de France, 8e édition, 1967, Bibliothèque de philosophie contemporaine, édition électronique réalisée à partir du livre d'Émile Durkheim (1897). Fichiers revus, corrigés avec ajout des mots grecs par Bertrand Gibier. Fichiers remplacés le 8 juillet 2008. http://classiques.uqac.ca/classiques/Durkheim_emile/division_du_travail/division_travail.html, consulté le 30 mai 2015.

Hazael-Massieux, M. C. (1989) « Formes brèves et chansons dans *Atipa* : la littérature orale dans la littérature écrite », *in* FAUQUENOY, M. (dir.), *Atipa revisité ou les itinéraires de Parépou*, PUCL/L'Harmattan, 1989.

Huygues-Belrose, V. « La musique en Guyane », *in La Grande encyclopédie de la Craïbe, Arts et traditions*, vol. 10.

Jolivet, M.-J. (1989) « Les formes de la tradition musicale en Guyane créole », Actes deu colloque *in Musique en Guyane* et Catalogue de l'exposition, 29 septembre – 25 novembre 1989, Bureau du Patrimoine Ethnologique et Conseil régional de Guyane, Cayenne.

Kwabena Nketia, J. H (1970) « Les langages musicaux de l'Afrique subsaharienne. Etude comparative », *La musique africaine*. Réunion de Yaoundé (Cameroun), 23-27 février 1970, organisée par l'unesco, *Revue Musicale,* n° 288-289, 36.

Lannoy, M. (1986) « L'instrument de musique, facteur et marqueur de l'identité culturelle chez les Sénoufo de Côte-d'Ivoire », *Vibrations*, ii, 49-59.

Picard, F. (2003) *La musique chinoise*, 2e édition, Paris, Editions YOU-FENG.

Pindard, M.-F. (2006) *Musique traditionnelle créole, le grajé de Guyane*, Matoury, Ibis Rouge.

Schafer, M. (1977) *Paysage sonore* (*The Tuning of the World*), Préfaces de Louis Dandrel et de Jean-Claude Risset, Postfaces de Christian Hugonnet et Nicolas Misdariis & Patrick Susini (Ircam), Editions Wildproject, coll. Domaine sauvage.

Sowande, F. (1970) « Le rôle de la musique dans la société africaine traditionnelle », *in La musique africaine. Réunion de Yaoundé (Cameroun) 23-27 février 1970, op. cit*., pp. 59-68, en particulier, 59-60.

Sully-Cally & Lezin (1990) *Musiques et danses afro-caraïbes*, Martinique, O Madiana Editions.

Valéry, P. (1944) *Variété V*, Paris, NrF, Gallimard, 87

MATH'N POP *VERSUS* MATH'N FOLK? A COMPUTATIONAL (ETHNO)MUSICOLOGICAL APPROACH

Mattia G. Bergomi
Music Representations Team,
IRCAM-CNRS-UPMC, Unimi, LIM
mattia.bergomi@ircam.fr

Moreno Andreatta
Music Representations Team,
IRCAM-CNRS-UPMC
moreno.andreatta@ircam.fr

1. INTRODUCTION

According to a programmatic article by Philip Tagg on theoretical, methodological and practical aspects of popular music studies (Tagg (1982)), folk or traditional music is one of the three possible kinds of music, together with classical (or art music) and popular music. This typology constitutes what Tagg calls an *axiomatic triangle* of musical genres, each of which being characterized by criteria such as the usual or unusual mass distribution, the existence of a circle of professionals or a circle of amateurs who produces and transmits it, the principle modality of storage and distribution (ranging from oral transmission, in the case of folk music, to the recorded sound, in the case of popular music), the anonymous versus authorial character of the underlying compositional process, and so on.

This typology, as well as any other traditional taxonomy, has been recently criticized by several scholars (see, e.g., Fabbri (2007)) stressing the existence of other music genres which are not included in Tagg's axiomatics (e.g. jazz music or the so-called *musiques actuelles*) and suggesting the necessity of substituting this typology with a finer taxonomy based on computational models focusing on musical objects and making use of different theoretical approaches (or methods) in order to carry on computer-aided music analysis. The real musical objects would not be musical pieces belonging to pre-established and universal taxonomies, but, as suggested by Franco Fabbri, "musical facts" and "musical events", each being characterized by a set of properties turning out to be associated into high-order families called *types* (See Fabbri (2007, 2014)).

Starting from this change of perspective in the study of musical genres, we discuss in this paper some possibilities of overpassing the main linguistic-based approach which seems to motivate Fabbri's original formulation of types classification - largely inspired by George Lakoff's and Mark Johnson's theory of metaphors (see Tendahl & Gibbs (2008)) to more mathematically-based formalizations of musical structures and processes. After recalling the principle construction of a computational linguistic-based approach aiming at symbolically notating the audio files of large data based of popular and folk music (Bimbot et al. (2012)), we show how a mathematical formalization of the harmonic system provides some additional structural criteria in order to apply the System & Contrast Model to symbolic music information retrieval and automatic genre classification. A new approach to genre classification has been discussed in Andreatta (2014).

2. COMPUTATIONAL APPROACH IN FOLK MUSIC ANALYSIS

This computational approach clearly shows the interest of approaching the question of types or genres or categories with respect to automatic classification and the necessity of developing generic tools, in particular for the analysis of the harmonic space. We will discuss the use of the *Tonnetz* and some possible generalizations (Bigo (2013)) as it has been carried on until now principally within the "Math'n Pop Project" (Andreatta (2015)). Some new interesting approaches in dynamic *Tonnetz* construction, as described in the ongoing doctoral thesis within Ircam's Music Representation Team (Bergomi (2015a)) clearly suggests the possibility of using dynamic *Tonnetze* in folksong research and automatic classification. The use of the computer-aided models not only provides more efficient computational methods of handling large amounts of information, but they definitively "help the scholar in achieving what he has been unable to achieve so far with non-technological means [by] taking into consideration an unlimited number of parameters and [...] finding correlations between each and every one of them" (Keller (1984)). Computational models enable one to detect and recognize patterns that, according to Sorce Keller, "might then be used in bringing together variants in an objective way", which would offer scholars the possibility "studying how musical traditions interact and influence each other and, which is putting it in another fashion, to what extent they are similar and therefore compatible" (Keller (1984)). This suggests interesting applications of the tools developed within the "Math'n Pop project" to the special field of computer-aided folk music analysis, as we will show by automatically analyzing the folk component of popular music databases (such as Quaero. See Figure 1).

2.1 An anisotropic geometrical approach to music analysis

The whole set of geometric tools for music analysis endowed the isotropic structure of the primal representation of pitch classes as $\mathbb{Z}/12\mathbb{Z}$ (see Andreatta (2015)). Generally, folk music has clear tonal or modal centers, declared either harmonically (triads, seventh chords or altered with

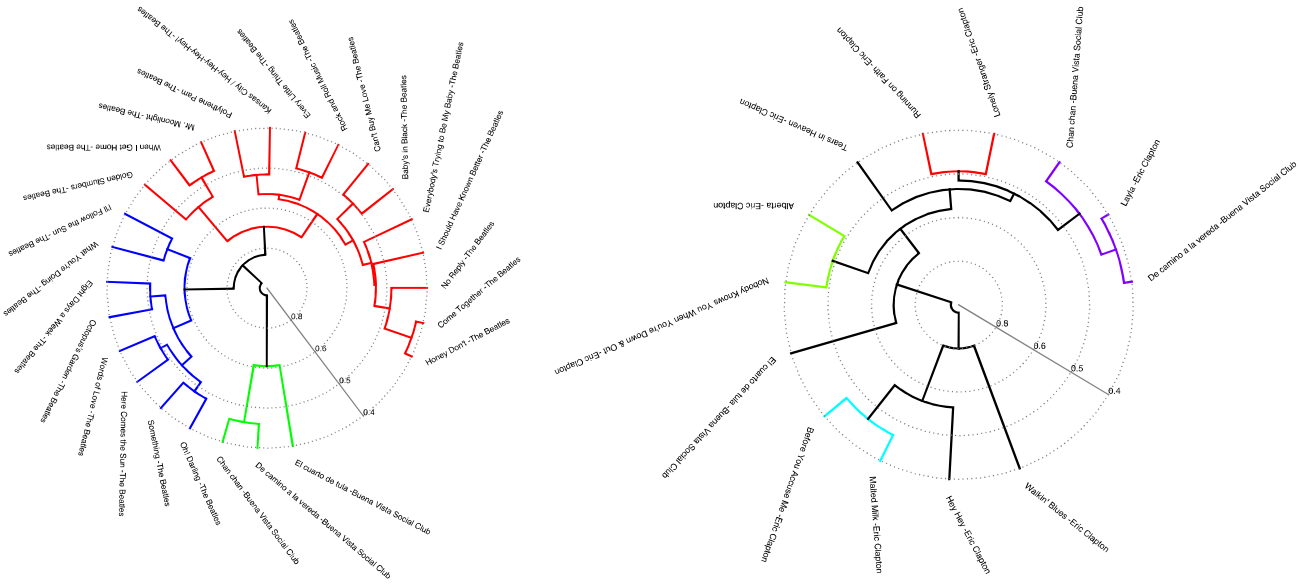


Figure 1: Two clusterings of Quaero database showing the place of the music by Buena Vista Social Club with respect to the Beatles and Eric Clapton's music respectively.

chord tones) and melodically (the voice generally stresses consonant notes of the arpeggio of the harmony, although it can be enriched and embellished with passage notes). This kind of feature suggest the idea to represent this kind of music introducing preferred directions. The model presented in Bergomi & Popoff (2015) allows to deform the standard geometric spaces used in music analysis, such as $\{(\mathbb{Z}/12\mathbb{Z})^n, \mathbb{T}^n/\mathcal{S}_n, \dots\}$ through a certain feature.

In particular, using the *Tonnetz* and the dissonance induced by the chord played in a certain time span Δt, it is possible to describe a song as a sequence of configurations of the *Tonnetz* representing either on an harmonic and melodic level the tension/resolution pattern in time. Each state is created via a deformation of the vertices of the *Tonnetz* (as shown in Figure 2).

2.2 Shapes classification

Each folk song can be represented as a sequence of chords and thus as a series of dynamical states, representing either a deformed configuration of the *Tonnetz* or a subcomplex. The method we suggest to characterize each shape obtained from the segmentation is Persistent Homology and it is borrowed from the the field of Computational Algebraic Topology [1]. Given a simplicial complex K, this particular strategy consists in rebuilding it through a filtration of the complex $\emptyset = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K$ and computing the homology for every subcomplex K_i for $i \in \{1, \dots, m\}$. The results of this kind of analysis can be represented as a *persistent diagram* which is a signature of each frame of the song.

The Wasserstein distance allows to compare persistence diagrams belonging to different shapes and hence to find related states belonging to a certain song, or to a set of songs. This method ha been discussed in Bergomi (2015b).

[1] See Edelsbrunner & Harer (2008); Zomorodian & Carlsson (2005); De Silva & Ghrist (2007); Carlsson et al. (2005); Cerri et al. (2006) and many others.

3. RESULTS, CONCLUSIONS AND FUTURE WORKS

If persistent homology is widely used in shape recognition and image analysis (see, e.g., Di Fabio & Landi (2011)), the application of these powerful algorithms to the symbolic/signal articulation is still very rare in the field of Music Information Retrieval. Applied to the domaine of popular music, and — more specifically — folk music, this approach gives a different point of view on the tonal and modal structure of existing songs. The aim of this analysis is twofold. On one side, comparing two surfaces in a 3-dimensional space is a task everyone can perform directly having an immediate intuition on the features stressed by the deformation of the surface; on the other side, the analysis of the geometrical properties of the shapes we computed enables one to classify and create explicit links concerning both the study of the inner structure of a song and its relation with other works. In particular, folk music, thanks to its structure, fulfills the hypothesis on the tonal (modal) nature of the music to be analysed.

We will end by suggesting some future research directions, in particular concerning the use of some other topological and algebraic approaches to folk music representation and classification. We are particularly interested in Formal Concept Analysis (Ganter & Wille (1998)), a paradigm which has been recently applied in music analysis by combining topological structures with ordered-structures (Freund et al. 2015). We will show how this approach enables one to open a wider discussion about the possibility of a general theory of classification (Parrochia & Neuville (2013)). If the question of the existence of universals in

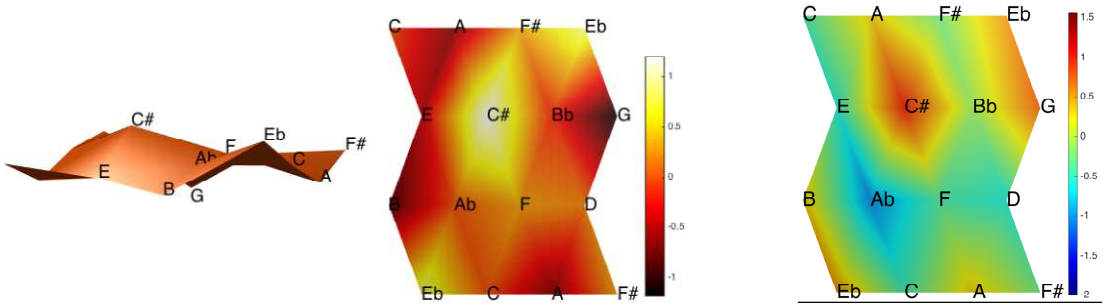


Figure 2: The *Tonnetz* whose vertices are deformed by the dissonance induced by a C_M chord, and its discrete Gaussian curvature.

music is still open in contemporary (ethno)musicology, the use of universal constructions in computational musicology has been definitively made possible thanks to the role played by algebraic, topological and ordered-structures representations and formalizations of musical events, independently of any existing underlying taxonomy. This would finally minimize the impact of ideology in disciplinary divisions, as recently suggested by Franco Fabbri in his critical overview of Music Taxonomies (Fabbri (2014)), and contribute to the emergence of new theoretical paradigms in the field of contemporary computational musicology.

4. REFERENCES

Andreatta, M. (2014). Math'n pop workshop: Formal and computational models in popular music. *ICMC/SMC Joint International Conference*.

Andreatta, M. (2015). Modèles formels dans et pour la musique pop, le jazz et la chanson. In *Esthétique & Complexité, Neurosciences, Philosophie et Art*. Editions du CNRS.

Bergomi, M. G. (2015a). Dynamical and topological tools for music analysis. *PhD Thesis, UPMC-Ircam-LIM. To be defended in December 2015*.

Bergomi, M. G. (2015b). Dynamics in modern music analysis. *Lecture delivered at the XXIst Oporto Meeting on Geometry, Topology and Physics.*

Bergomi, M. G. & Popoff, A. (2015). Consonance-based structuration of musical entities. *Journal of Mathematics and Music (submitted)*.

Bigo, L. (2013). Représentations musicales symboliques à l'aide du calcul spatial. *PhD Thesis, Université Creteil / Ircam*.

Bimbot, F., Deruty, E., Sargent, G., & Vincent, E. (2012). Semiotic structure labeling of music pieces: concepts, methods and annotation conventions. In *13th International Society for Music Information Retrieval Conference (ISMIR)*.

Carlsson, G., Zomorodian, A., Collins, A., & Guibas, L. J. (2005). Persistence barcodes for shapes. *International Journal of Shape Modeling*, *11*(02), 149–187.

Cerri, A., Ferri, M., & Giorgi, D. (2006). Retrieval of trademark images by means of size functions. *Graphical Models*, *68*(5), 451–471.

De Silva, V. & Ghrist, R. (2007). Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, *7*(1), 339–358.

Di Fabio, B. & Landi, C. (2011). A mayer–vietoris formula for persistent homology with an application to shape recognition in the presence of occlusions. *Foundations of Computational Mathematics*, *11*(5), 499–527.

Edelsbrunner, H. & Harer, J. (2008). Persistent homology-a survey. *Contemporary mathematics*, *453*, 257–282.

Fabbri, F. (2007). The king is naked: The musicological unified field and its articulation. *British Forum for Ethnomusicology Annual Conference 2007*.

Fabbri, F. (2014). Music taxonomies: an overview. *JAM 2014,"Musique savant / musiques actuelles: articulations", Ircam, 15-16 December*.

Ganter, B. & Wille, R. (1998). *Formal Concept Analysis: Mathematical Foundations*. Springer.

Keller, M. S. (1984). The problem of classification in folksong research: a short history. *Folklore*, *95*(1), 100–104.

Parrochia, D. & Neuville, P. (2013). *Towards a general theory of classifications*. Springer.

Tagg, P. (1982). Analysing popular music: theory, method and practice. *Popular music*, *2*, 37–67.

Tendahl, M. & Gibbs, R. W. (2008). Complementary perspectives on metaphor: Cognitive linguistics and relevance theory. *Journal of Pragmatics*, *40*(11), 1823–1864.

Zomorodian, A. & Carlsson, G. (2005). Computing persistent homology. *Discrete & Computational Geometry*, *33*(2), 249–274.

The evolution of the *takht* in new acoustic environments: cycle of evolution and acoustical indices

Yosr Bouhali[1, 2]
yosr.bouhali@gmail.com
[1]Sorbonne Universités, Univ. Paris-Sorbonne, IREMUS

Jean-Dominique Polack[2]
jean-dominique.polack@upmc.fr
[2]Sorbonne Universités, UPMC Univ. Paris 6, Institut d'Alembert

ABSTRACT

The typical formation of the *takht* that we find up until the first half of the XXth century is constituted by five musicians like all other formations of the same period. This theoretical plan is subject to numerous variations. We meet for example a violin which replaces the *rabâb* for the interpretation of *basheref*[1] and *ashrâl*[2]. The harmonium or the piano are also introduced to the formation at the beginning of the XXth century which then takes the shape of a sixfold *takht*.

The introduction of the small formation in new acoustical environments accelerates its evolution. In fact, this phenomenon is piloted by the multiplication and the deployment of instruments and singers on the stage, and by the diversification of the big formation in particular. We propose indicators for the modified formations in the new acoustical environments through the analysis of photos and newspaper articles. The continuous sound emission is articulated around the sound sources in the acoustical environment, which also constitute the articulation of the formation: the spatial disposition of the musicians. Our goal is to presents phases, which model the cycle of evolution of the *takht*.

We also analyse the room-acoustical indices of eight concerts venues representing the typical architectural spaces where the traditional Tunisian music is played. In fact, we compare the acoustics of the music room of *Ennejma Ezzahra* Palace (Baron d'Erlanger's Palace) to that of Sadok Bey's palace (Bardo museum), the Café Mourâbit, Dâr Rachidia, Zaouïa of Sidi Belhassen, Tunis Municipal Theatre, Carthage Theatre, and El-Jem Amphitheatre. In order to characterize and evaluate the transfer to the new acoustical environments, we analyse these new environments through their abilities to enhance the execution of the *takht* ensemble.

1. INTRODUCTION

The *takht's* formation interprets the traditional Tunisian music commonly known as *ma'lūf*. In order to characterize its recent metamorphosis,We propose to define its cycle of evolution by studying photos of the formations of the Rachidia's Institute successively over the years. We define indicators for the modified formations pin the new acoustical environments. We should point out that the *ma'lūf's* most valuable asset is the fact that it had been systematically studied, preserved, promoted and presented by this institute. After Independence, the *ma'lūf* was designated, as National Musical Heritage and the Rachidia became the model and source for *ma'lūf's* practices throughout the nation. The government's contribution only consisted in transposing the Rachidia's method and achievements into its cultural and educational program. It extended the Rachidia's effort to spread the *ma'lūf* from their centre in Tunis throughout Tunisia[3].
Thus, the new acoustical environments influence the evolution of this small formation. Indeed, it grows and opens up as it starts playing in these new environments. Therefore, we propose to study the room-acoustical indices of eight iconic concert venues in order to evaluate their abilities to enhance the execution of the *takht* ensemble, including the latest versions of this formation.

2. *TAKHT* EVOLUTION CYCLE

We define the cycle of evolution of the *takht* as being the path it followed when introduced in large acoustical spaces. The most important event of the whole cycle corresponds to its introduction within new spatial frames. In the first phase, the small formation multiplied the number of instruments as shown by a 1935 photo of the *Rachidia* ensemble, bringing together for the first time 24 musicians at their first concert in Tunis Municipal Theatre. Six violins, five *oud*, a *rabâb*, three *qânûn*, a player of *naqqarât*, an other one for the târ and seven singers respond to the call of the institution and rehearsed for one year for this particular event. The 1937 photo of the ensemble in Tunis Municipal Theatre reveals the frame of the production and the disposition of the musicians on stage. A single female singer takes position in the middle of the chorus-singers in the first row. On July 24th, 1972, the Rachidia ensemble performs for the first time on the stage of the Roman theatre of Carthage. The Tunisian radio & television team as well as the Stars of Tomorrow band merged for the event.

[1] Instrumental form used in the progress of a concert of Tunisian classic songs, generally at the beginning of a musical suite.
[2] Plural of *Shghul*: it is a set of songs of the foreign kind "*Tawshih* " in the classic repertory of *nawba* suite, one of the main forms of Tunisian classical music.
[3] Ruth F. Davis, *Ma'lūf: Reflections on the Arab Andalusian Music of Tunisia,* Lanham (*Md*), Scarecrow press, 2004, *p.51-71*.

Figure 1: The first Rachidia ensemble with his director Mostapha Sfar placed in the middle, in 1935.
All the musicians are men apart from the soloist singer Hassiba Rochdi.[1]

Figure 2: The Rachidia ensemble at the Tunis Municipal Theatre in 1937.[2]
The musicians are arranged in two semicircles, displaying from the outside inward: the instruments; then the choir with the soloist singer always placed in the middle.

[1] Mohamed Sakanjî, *Al-Rachidia madrasatou Almûsikâ wa alghinâ' alarabî fî tûnis* [The Rachidia is a school of music and Arabic singing in Tunisia], Tûnis, Chérikét' Kâhiyā linéchr, 1986, p105.
[2] *Ibidem.*

Figure 3: The *Rachidia* ensemble, composed by more than 90 musicians, in its first appearance on the Roman theatre of Carthage on July 24th, 1972.[3]

In the second phase of its cycle, the formation evolves as can be seen on the photo of Rachidia taken in 1940. Even if we cannot accurately count the number of musicians, due to the poor quality of the photo, the ensemble clearly is much broader and exceeds 29 musicians, in comparison to the 1935 photo and the one taken at its headquarters with 29 musicians in 1939. The transition from a purely masculine choir with a female soloist singer, to a mixed choir consisting of five women and as many men, is another feature of this second phase.

Figure 4: The Rachidia ensemble at Tunis Municipal Theatre in 1940.[4]
In the first semicircle, the mixed choir composed of five women is placed between five men and a conductor to manage the formation.

The third phase relates to the evolution of the ensemble towards the end of the fifties. The arrangement of the musicians on three rows on a 1957 photo and the choir formed by twenty-two members, among whom eight women and fourteen men, on another photo from the late fifties, emphasizes this expansion.

[3] Salah Mehdi, *maqâmêt Elmûsîqâ alarabiyya* [The modes of arabic music], Tounis: Echerikā altounisiyā lifûnûn alrasm, 1982, p155.
[4] *Ibid.*, p.106.

Figure 5: Mohamed Triki conducts the Rachidia ensemble during a show at Tunis Municipal Theatre in the late 50's.[1]
A strengthened choir, arranged in three rows to the left of the instrumentalists, with 8 women in the first row, 9 in the second and 5 in third.

Figure 6: The Rachidia ensemble conducted by Abdelhamid Ben Aljia in 1957.[2]
On this photo, the musicians occupy a wide side area and are arranged on three rows, with use of a sound system. Notice the *nây* position next to the *oud*.

In the next phase, we observe the same instrumental evolution with 36 musicians or so in 1967. In 1978, the number of musicians performing on the stage of the Roman theatre of Carthage reaches 41. This builds the fifth phase.

Figure 7: A photo of Rachidia at Tunis Municipal Theatre in 1967, conducted by Salah Mehdi[3].
36 musicians appear on the photo. The choir stands behind the instruments on this photo. Musicians' arcs concentrate more.

Figure 8: The Rachidia ensemble at the Roman theatre of Carthage in 1978 formed by 42 musicians among whom 25 chorus-singers.[4]

In 1981, the ensemble reaches its sixth phase, with a smaller orchestra of 27 musicians and a choir of 18 members. A photo of the ensemble in 1985 gives evidence that the ensemble gets even bigger than in 1978 even if the picture does not allow to count all the musicians.The seventh phase corresponds to more sound expansion, with a choir that now reaches 26 members in addition to the multiplication of the cellists.

[1] Photo taken from a prospectus written for the opening of the 2008-2009 session and the commemoration of the 50th anniversary of the death of the Tunisian singer Saliha, p.8
[2] *Ibid.*, p.109.

[3] Photo supplied by Taoufik Ben Khlifa, archivist and member of the choir of the *Rachidia*, during an interview on 05-03-2015.
[4]Capture sreen of a video broadcasted on https://www.youtube.com/watch?v=WTB-U-I4oaQ&list=PLEJjgui6-IUadcWppPgAdORvqKJ5h_Pv8 on 28-05- 2015.

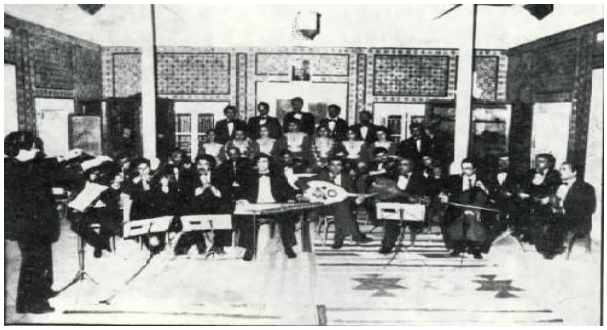


Figure 9: The Rachidia ensemble conducted by Lotfi Loukil in 1984.[1]
Better spatial organization. The musicians are more separated to give better visibility to the listener. Percussion instruments are placed next to the columns. The choir, placed between columns, stands up on one row, with the highest point corresponding to the tallest singers. Conductor position in diagonal with regard to percussionists in order to clear view for listeners. *The arrangement of the ensemble is the same as in 2006 (Figure 11).*

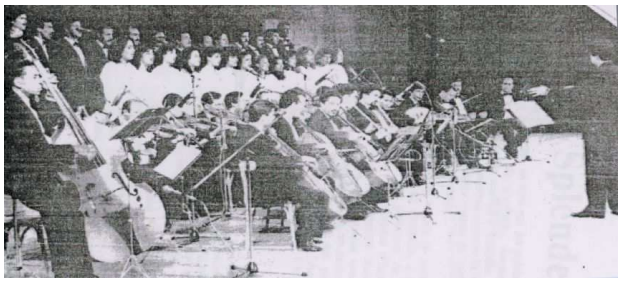


Figure 10: The Rachidia ensemble in 1985.[2]
Influence of modernization is visible in the increase of the number of cellos and a choir of 26 members. The presence of microphones formalizes the evolution of the sound reinforcement at this period.

In the 90s and in the early 2000s, the ensemble combines its highest number of instruments and reaches its eighth phase that branches out in three axes. In the first one, the ensemble with now 45 musicians conducted by Abelhamid Belalgia, calls on the musicians from the Tunisian Radio station for back up. It is characterized by a richer and wider sound. In the following years, the ensemble, under the direction of Zied Gharsa, iconic figure of the *ma'lūf* in Tunisia, counts the same number of musicians when it performs in Tunis Municipal Theatre. Some differences do exist, however, with regard to the organization and the diversification of the instruments. One distinguishes the "double" formation, consisting in pairs of instruments whose multiplication does not stifle the presence of traditional instruments of the *takht*. The second axis leads to supernumerary formations with more than 100 artists present on stage of the Roman theater of Carthage as we can see on the photos of the Rachidia ensemble in the Roman theatre of Carthage in 2007 and 2012. The arrangement of the singers and instruments is similar to the one of the 30s, 40s, and 1957 in Tunis Municipal Theatre. In the third axis, the formation resumes to the same number of musicians as in its first appearance in Tunis Municipal Theatre in 1935, and thus returns to its starting point at the beginning of the cycle. This step was prepared by the second axis and it allows a glimpse on the appearance of the supernumerary formation, which is similar to one that took place in 1972 in its first appearance on the Roman theater of Carthage. The continuity that runs from the first to the final version of the companies is transparent despite the metamorphosis of the ensemble.

Figure 11: The Rachidia ensemble in 2006 in Tunis Municipal Theatre.[3]

42 musicians combined in pairs, hence the name of "double" formation. The *oud* outstrips the *qânûn* and positions in the middle of the stage to guide the orchestra.

[1] Mohamed Skanjî, *op.cit.*, p.118.
[2] A.M, (July 1985), *"La Rachidia à Carthage comme une renaissance"* [Rachidia in Carthage as a revival], La Presse (Tunis), p.9.
[3] Photo on cover of a prospectus distributed to the public on the occasion of the opening concert of the cultural session on December 8th, 2006 in memory of Mohamed Triki.

Figure 12: Rachidia's troop during the closing evening of the International Festival of Carthage in 2007.[1]

The ensemble opens up its instrumental sections and spreads on the whole stage area The arrangement of choir singers and singer soloists in front of the scene, between the traditional instruments, calls back to the 30s, 40s and 1957 as show in photos above. Two choirs join efforts in response to the soloist, creating a huge sound upsurge.

Figure 13: The Zied Gharsa ensemble at the opening of the International Festival of Carthage in 2012[2]
A widely supernumerary formation, constituted of more than 100 artists, is gathered for this concert.

The singers arrange according to the same choir and singers model as for the closing concert of Rachidia in 2007.

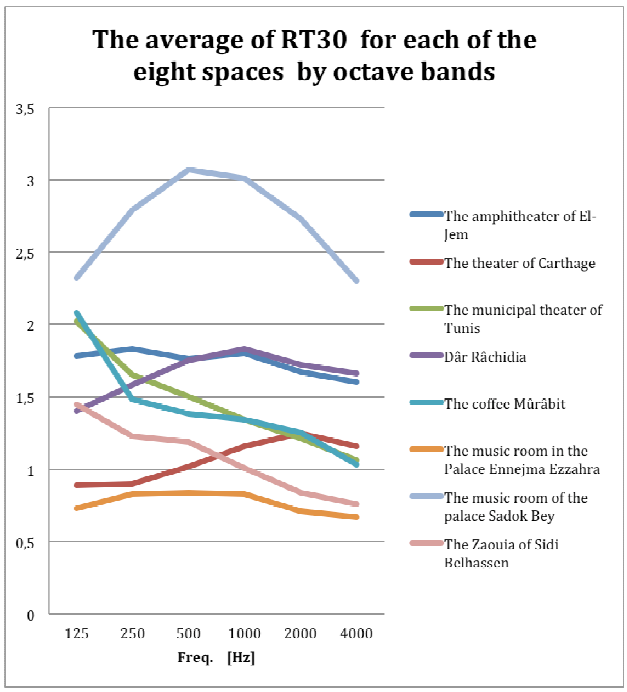
Figure 14: The Zied Gharsa ensemble formed by 24 musicians on the Roman theater of Carthage in 2013[3]

A formation, which is identical by its number to the one, which occurred on the municipal theater of Tunis for the first time in 1935. The ensemble is arranged in two curved rows where the two singer soloists are valorized by projectors.

3. STUDY OF EIGHT TYPES OF ROOMS: OBJECTIVE PARAMETERS VS. *TAKHT*'S REQUIREMENTS

We compare the parameters acoustics of the music room of *Ennejma Ezzahra* Palace with the seven other spaces. Indeed, it is the archetypal place of musical research and exercises of the musicians of the formation of *takht* which participated in the 1932 Congress in Cairo.

3.1. Reverberation Time RT30



[1] Capture screen of a video broadcasted on YouTube https: // www.youtube.com / watch? V=bVWkTAfDEPI consulted on1 6-04-2015.

[2] Capture screen of a video broadcasted on YouTube https://www.youtube.com/watch?v=Xn4BnFOiprs consulted on 19-04-2015.

[3] Capture screen of a video broadcasted on YouTube https://www.youtube.com/watch?v=zbxPUHCeshA consulted on 31-05-2015.

The reverberation times (RT30) of the Zaouia of Sidi Belhassen and of Carthage Theatre are closest to that of the music room of *Ennejma Ezzahra* Palace, a small square room of 4 m side and a height bordering 5 m. Reverberation in Carthage Theatre is almost symmetrical to that of the Zaouia. The three are classified together in the range 0.6 - 1.5 s. These values are recommended for lecture halls or small theatres. The Zaouia of Sidi Belhassen whose dimensions equal 4 x 15 x 6. 35 m seems to be closer to hold a more intimate music than the Roman Theatre of Carthage whose *cavea* diameter measures 105 m and its depth exceeds 25 m.

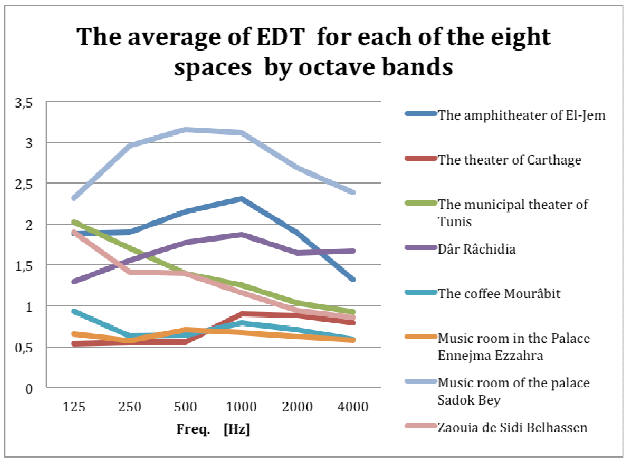
The reverberation times in the other areas differ even more from the RT30 of the music room of *Ennejma Ezzahra* Palace.

The curve of RT30 of the Tunis Municipal Theatre and Café Mûrâbit takes a shape of similar decline with a reverberation time between 2.1 s to 1 s with frequency. The ascent of reverberation time in the bass frequencies compared to the middles ones in the great hall of Tunis Municipal Theater, which exceeds 20 m in height and a capacity rising beyond 1350 people, is desirable. The reduction in the reverberation time on the 2kHz octave band avoids sharp and aggressive sound in the theatre. The Mûrâbit Café, an 8 x 12.5 x 5.6 m room with a 8.4 m long and with an entrance hallway dimensioning 8 x 2.8 m is very reverberant for traditional Tunisian music interpreted by the *takht*.

The reverberation time of Dâr Rachidia which measures 16.4 x 13 x 8.4 m is close to that of El-Jem amphitheatre, despite that the first is a small volume with a covered courtyard, whereas the second is a 36 m high amphitheatre with arena axes measuring 64.5 m and 38.8 m respectively. The last one shows a constant reverberation time throughout frequencies, thanks to its highly reflective construction material, namely the freestone...

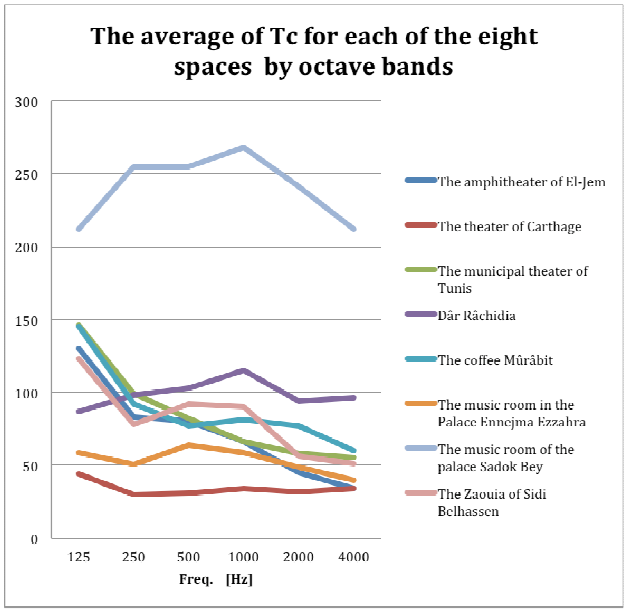
The RT30 of the music room of the Sadok Bey Palace, of dimensions that are equal to 24 x 6.8 x 9 m, is very high, but it certainly does not correspond to what it was at the time of beys, because of the absence of carpets and the presence of numerous Roman mosaics on the walls and floor.

3.2. Early Decay Time EDT



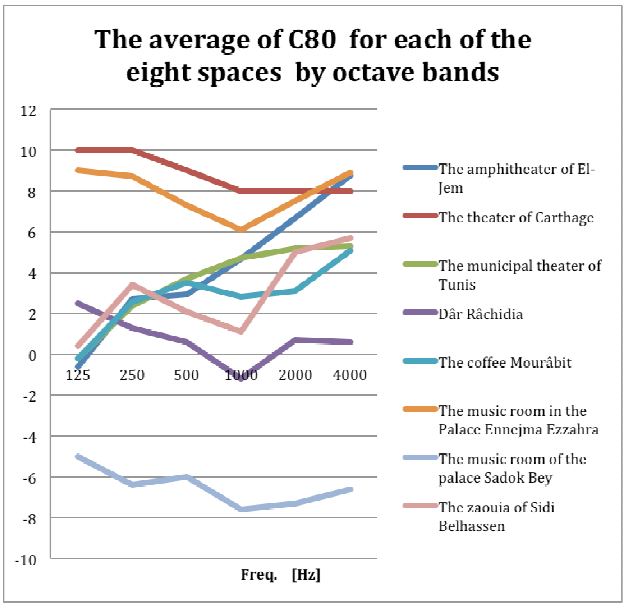
The Café is very similar to the music room in the Palace *Ennejma Ezzahra* regarding the early decay time EDT, which corresponds to the perceived reverberation. However, the sound in the Café is louder than in the Palace of the Baron, as will be shown below. These two areas allow a priori to appreciate the dynamics and the attacks on traditional instruments that form the *takht*.

3.3. Centre Time Tc



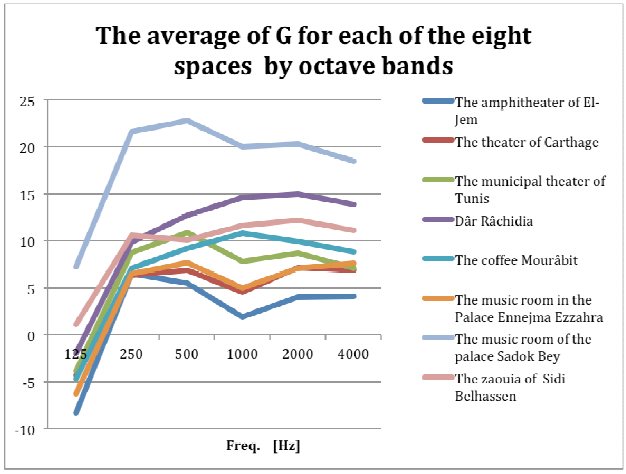
The Theatre of Carthage's Tc is the closest to that of the music room of Palace *Ennejma Ezzahra*. Besides, Tc is suitable for its small volume but low for the large Carthage Theatre. Tc is also very long for Tunis Municipal Theatre, Zaouia of Sidi Belhassen and Café Mûrâbit. It is much too long in El-Jem Amphitheatre and Dâr Râchidia, where they are not conductive to good appreciation of a *nawba* interpreted by the traditional *takht*.

3.4. Clarity C80



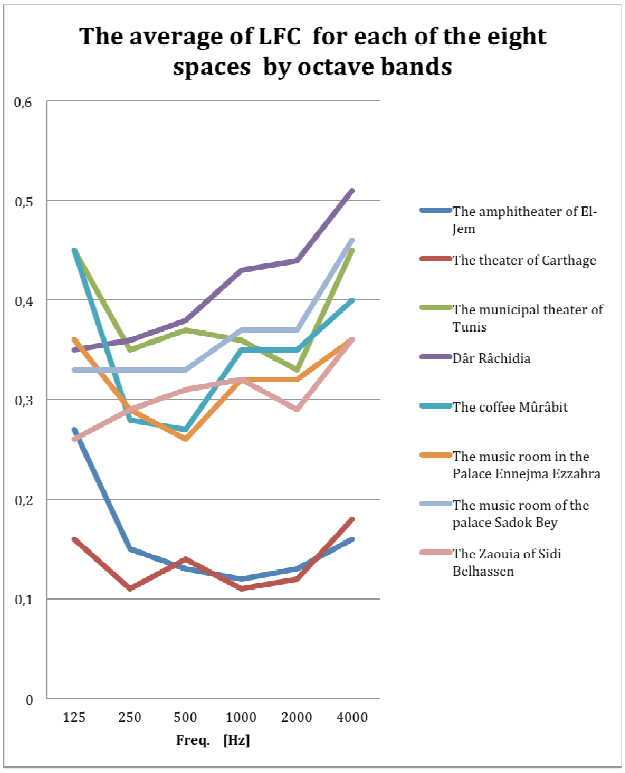
The values of clarity C80 at middle frequency allow to group the various spaces as follows:
- The music room of Sadok Bey Palace (Bardo Museum), with its clarity C80 lower than 6 dB, is a very reverberant and, in a way, confusing room because the direct sound and the first reflections will be drowned in the late reflections. The Zaouia of Sidi Belhassen and Dâr Râchidia have an indication of clarity in the range 1.2 - 2.1 dB, close to what is recommended for chamber music.
- Café Mûrâbit, Tunis Municipal Theatre and El-Jem amphitheatre have clarity in the range 2.9 - 4.7 dB. Generally, this valuable lapse allows for the appreciation of the sound attacks and dynamics. They are recommended for the appreciation of pop or rock music.
- Clarity in the music room of the *Ennejma Ezzahra* Palace and Carthage theatre in the octave bands centred on 500 and 1000 Hz are between 6.1 and 9 dB. These values are usually considered as detrimental to the appreciation of the music because there will be very few reflections and the sound will be very dry. But it allows to hear all the subtleties of the music and dry environments will enhance the *takht*.
We deduce from this analysis of clarity that Café Mûrâbit would second best suit the *takht*, due to the fact that it has the second smallest volume after the music room of the Erlanger Palace. Optimal C80 would be in the range 2.8 - 3.5dB.

3.5. Strength G



Sound strength G has to be in the range 2 – 10 dB according to ISO 3382 standard. The curves shows that Dâr Rachidia and the music room of Sadok Bey Palace exceed by far the values recommended for this index.
Indeed, the strength of Dâr Râchidia ranges from 12.7 to 14.7 dB, and from 20 to 22.8 dB in the music room of Sadok Bey Palace on octave bands of 500 and 1000 Hz.
The theatre of Carthage and *Ennejma Ezzahra* Palace music room have strengths that fell in the range recommended by the standard. The sound volume of the *takht* will nevertheless be somewhat weak in the theatre.
The Municipal Theatre of Tunis, the Mûrâbit Café, and Zaouia of Sidi Belhassen are neighbouring the recommended values. The theatre will not, for sure, amplify the kind of formation in question. The Café and the Zaouia are thus stronger from the point of view of sound volume compared to the strength of the music room of the palace *Ennejma Ezzahra*.
El Jem amphitheatre has the lowest values compared to the music room of the palace *Ennejma Ezzahra*. They are close to the recommended values, with Strength from 1.9 to 5.5 dB in the octave bands 500 Hz and 1000 Hz.

3.6. Lateral Energy LFC



The Zaouia of Sidi Belhassen has LFC closest to the music room of the Palace of *Ennejma Ezzahra*. All the values of LFC are lower than 0.36, within the range 0.26-0.36. The amphitheatre of El-Jem and the theatre of Carthage have an average LFC lower than all the other spaces, in the range 0.12 - 0.18 except for the 125 Hz octave band in El-Jem. All these values fall within the range of 0.05 - 0.35 recommended in the ISO 3382 standard.
These ranges will result in a broad sound in the former venues, but a narrow one in El-Jem and Carthage.

Figure 15: Tunisian *takht* participating in the Congress of Cairo in 1932[1]
From left to right: singer, *târ*, *rabâb*, *oud* and *naqqarât*

[1] Mohamed Garfi, *Les formes instrumentales dans la musique classique de Tunisie* [The instrumental forms in the classical music of Tunisia], Tunis: Sotepa Graphic, 1996, p.64.

4. CONCLUSION

The music room of *Ennejma Ezzahra* Palace, thanks to its small volume and its values of the room-acoustical indices, suits the intimate character of the music played by the *takht* which has evolved as it became played in large spaces. This assumes that all the Tunisian traditional houses, which are defined by the same room disposition, promote listening to the Tunisian traditional music at the time. The *takht* life cycle allows understanding the links, which connect the current formations with those they preceded it. While keeping trace of the original formation, they were gradually transformed into another type, with multiplication of instruments and voices. The new acoustical environments are better suited to these new formations, which are more sonorous and resounding and, at the same time, explain the transition that has taken place in them.

5. REFERENCES

A.M. (July 1985). *"La Rachidia à Carthage comme une renaissance" [The Rachidia in Carthage as a revival]. La Presse* (Tunis).

Beranek. Leo Leroy (2004). Concert halls and opera houses. New York : Springer.

Davis. Ruth F. (2004). *Ma'lūf: Reflections on the Arab Andalusian Music of Tunisia,* Lanham (Md): Scarecrow press.

Garfi. Mohamed. (1996). Les formes instrumentales dans la musique classique de Tunisie [The instrumental forms in the classical music of Tunisia], Tunis: Sotepa Graphic.

Guettat. Mahmoud. (2000). La musique arabo-andalouse [The Arabo-Andalusian music], Paris: El Ouns.

ISO 3382. (2009). Acoustics-Measurements of room acoustics parameters.

Kuttruff. Heinrich (1991). Room acoustics. New York: Elsevier Applied Science.

Mehdi. Salah. (1982). *Maqâmêt Elmûsîqâ alarabiyya* [The modes of arabic music]: Tounis: Echerikā al-tounisiyā lifûnûn alrasm.

Mehdi. Salah. & Mohamed. Marzouhi. (1981). El Mahad alrachîdî lilmousîkâ at-tounisiyyā [The Rachidia's institute of Tunisian music], Tounis: Echerikā altounisiyā lifûnûn alrasm,

Sakanjî. Mohamed. (1986). *Al-Rachidia madrasatou Almûsikâ wa alghinâ' alarabî fî tûnis* [The Rachidia is a school of music and Arabic singing in Tunisia], Tûnis, Chérikét' Kâhiyā linéchr.

THE HUKWE BOW SONG OF THE SYMPOSIUM ON TRANSCRIPTION (MIDDLETOWN, 1963) FIFTY YEARS LATER: NEW PERSPECTIVES, METHODOLOGIES AND ANALYSES

Paolo Bravi
Conservatorio di musica "G. P. Palestrina" - Cagliari
pa.bravi@tiscali.it

Cécile Delétré
IreMus -Université Paris-Sorbonne
cecile.deletre@gmail.com

Marco Lutzu
Università "Ca' Foscari" - Venezia
mlutzu@livestudio.it

Emmanuelle Olivier
Centre Georg Simmel
UMR EHESS-CNRS 8131
olivier@ehess.fr

François Picard
IreMus -Université Paris-Sorbonne
Francois.Picard@paris-sorbonne.fr

Alice Tacaille
IreMus -Université Paris-Sorbonne
al_tac@club-internet.fr

1. INTRODUCTION

On November 2, 1963, at the Eight Annual Meeting of the Society for Ethnomusicology at Wesleyan University, a session was dedicated to a "Colloquium on Transcription and analysis" from which, almost one year later, derived a long paper appeared on the journal Ethnomusicology under the title "Symposium on Transcription and Analysis: a Hukwe Song with Musical Bow", written by Nicholas England with contribution by Robert Garfias, Mieczyslaw Kolinski, George List, Willard Rhodes and moderated by Charles Seeger (England, et al., 1964). The object of the Colloquium and of the subsequent paper was a recording of «'Du:», a song performed by an old man of the [ΣιδΖΕλΕ] community, Kafulo, with the accompaniment of a musical bow, and recorded by England during a short exploratory visit at the bushman group in Namibia (at that time South West Africa) four years before. The song was transcribed and analysed by four ethnomusicologist – Robert Garfias, Mieczyslaw Kolinski, George List, Willard Rhodes – who worked independently from each other and without any specific knowledge of the field, of the musical culture and of the performer, as well as of the lyrics and meaning of the song. The experiment had a great success and received a strong attention on practically every textbook of ethnomusicology, as example of the limits of the staff notation for the transcription of non-western music and subjectivity of the musical interpretation (Agamennone, Facci, Giannattasio, & Giuriati, 1991; Magrini, 2002; Giannattasio, 1998; Càmara de Landa, 2003). In particular, the synoptic view proposed by Charles Seeger (Figure 1) was a "must" in all discussion on the methodology of representation and analysis of (ethno)musical forms.

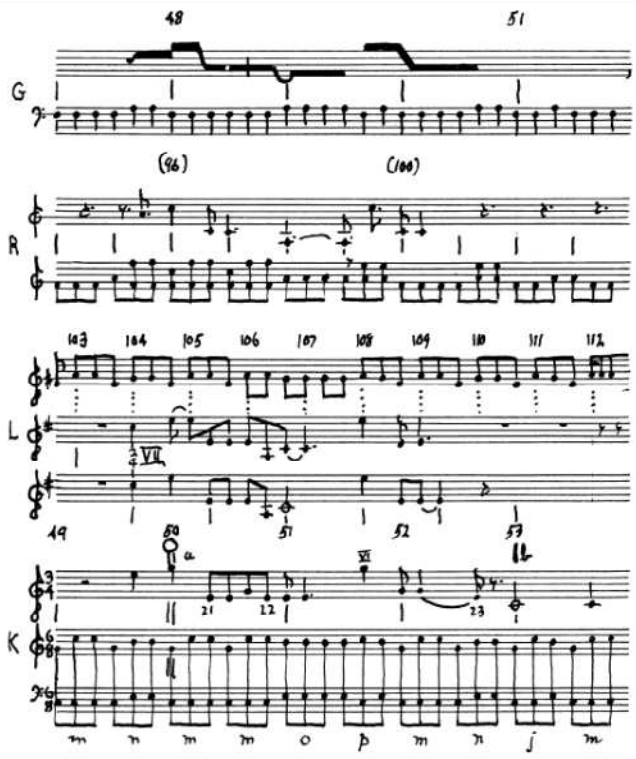


Figure 1. The first page of the synoptic view with the four transcriptions of Garfias (first staff, labelled "G"), Rhodes (second staff, "R"), List (third staff, "L"), Kolinski (fourth staff, "K").

The aim of the work is to propose new perspectives and insights on the recorded song, based on the evidence that the revolution in the methodologies of research and representation of sound may allow for accurate and efficient musical analysis, and that the cultural background and musical expertise of the transcribers can bring them to give different interpretation of the structure of the song. In this case, six of them – Cécile Delétré [from now on: CD], Marco Lutzu [ML], Emmanuelle Olivier [EO], François Picard [FP], Alice Tacaille [AT] – were engaged to transcribe and/or analyze the song, and one – Paolo Bravi – had the role to coordinate the work and to com-

pare and synthesize the results of their analyses, taking also into consideration the transcriptions and analyses proposed by the ethnomusicologists in the cited in the well-known 1964 paper.

2. MATERIAL AND INSTRUMENTS

Musical transcription and analyses have dramatically changed and improved its potentials with the so-called digital revolution. In this case, the original recording was given to the musicologists in a digitized form from the original published disk (frequency of sampling: 96 kHz, bitrate: 16 bits). Most of the musicologists who participated in this experiment of multiple transcription and analysis used specific software for both musical transcription (*Finale* [CD, FP] - Finale, 2014), listening – with possible slowering – annotation and different kinds of sound analysis (*Audacity* [CD, FP] - Audacity, 2014), *Sonic Visualizer* [ML, FP] - (Sonic Visualizer, 2014), *Praat* [CD, FP] - Boersma & Weenink, 2014) and synchronisation and visual representation of sound (*Acousmographe* - Acousmographe, 2014, *iAnalyse* - iAnalise4, 2014).

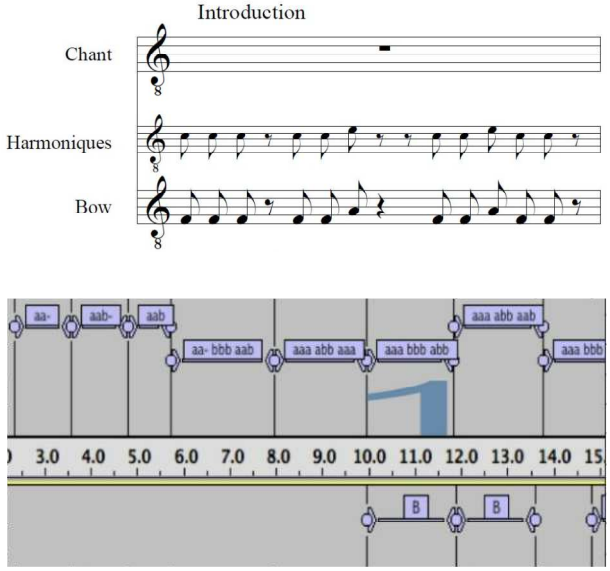


Figure 3. An example of transcription: the introduction of the song as reported and written down through Finale on a three staves staff (voice, harmonics of the bow, fundamental note of the bow) by CD [high]; an example of annotation: the first part of the song as annotated through Audacity (top panel: bow; bottom panel: voice) by FP [low].

3. ANALYSIS AND DISCUSSION

Analysis carried out by the musicologists regard different aspects of the musical structure, as form, rhythm, scale, melodic patterns, relation voice/accompaniment, spectral characteristics. Analyses in all these fields show both commonalities and different evaluation of specific aspects by different musicologists.

One of the most relevant of this aspects regards rhythm, which appears to be relatively regular as far as the bow is concerned – fundamentally based on grouping of three beats, but defined either as "an ostinato with variations, on a strict 9/8 meter" [FP] or "a melodic cycle of 18 quavers" [CD] – and in free rhythm (but with some regular pattern) as far as voice is concerned. However, different interpretations are presented regarding the starting point of the bow pattern (a cycle of three lowest notes for FP; a cycle of three highest notes for CD).

Another relevant aspect regards the voice. Despite the absence of a transcription and translation of the text, some musicologists noted that the same verbal expression (*judi* or *djoli*) is used at the end of each melodic phrase. This use of a recurrent textual element as a frontier to delimit the boundaries of musical phrases represent a new approach with respect to the analysis proposed in 1964.

Reading and comparing the analytical work of the five musicologists, which come after that of the four transcribers/analysts made in 1963, show that the evolution in the tools used in musicology allows a fine-grained description and analysis of many aspects of a musical performance which a simple listening, however skilled as it might, cannot afford. Nonetheless, it also emerges that there is a gap between what instruments measure and what the ear of a particular subject recognize as musically perceivable and relevant. In this respect, the new analyses on the Hukwe song confirm what was observed by the musicologist in the 1963 symposium: some consider as musically relevant the upper partials of the bow sound, and transcribe the main harmonic hearable above the fundamental (CD, as L, K, R in 1963, with clear differences among the notation of this sound); others say that harmonics are not clearly distinct and do not notate them (AT and FP, as G in 1963). This confirms that a strong degree of subjectivity may enter in the perception, musical evaluation and transcription of a song, particularly when it does not come from modern western musical cultures. But instrumental analysis may induce to similar issues if different procedures of measuring are applied.

The experiment also proved that there are heuristics that, after various decades, are still in use and considered as essential tools of the trade of the XXIth century musicologists, beyond technological advancements. The paradigmatic method of analysis, based on the idea of representing melodies in order to allow an immediate evaluation of the repetitions and variations within a musical piece, was adopted in this case by various musicologists (AT, CD, FP) (Ruwet, 1972). On the other hand, one can observe that this kind of approach can be applied at different levels and therefore different aspects of the musical structure of the song may appear and different interpretations may be put forward.

4. REFERENCES

Acousmographe. (2014). http://www.inagrm.com/accueil/outils/acousmographe

Agamennone, M., Facci, S., Giannattasio, F., & Giuriati, G. (1991). *Grammatica della musica etnica.* Roma: Bulzoni.

Audacity. (2014). Retrieved from http://audacity.sourceforge.net/?lang=it

Boersma, P., & Weenink, D. (2014). *Praat: doing Phonetics by computer*. Retrieved from http://www.fon.hum.uva.nl/praat/

Càmara de Landa, E. (2003). *Etnomusicologìa.* Madrid: Ediciones del Instituto Complutense de Ciencias Musicales.

England, N. M., Garfias, R., Kolinski, M., List, G., Rhodes, W., & Seeger, C. (1964). Symposium on Transcription and Analysis: A Hukwe Song with Musical Bow. *Ethnomusicology, 8*(3), 223-277.

Finale. (2014). http://www.finalemusic.com/

Giannattasio, F. (1998). *Il concetto di musica. Contributi e prospettive della ricerca etnomusicologica.* Roma: Bulzoni.

iAnalise4. (2014). http://logiciels.pierrecouprie.fr/?page_id=672

Magrini, T. (2002). *Universi sonori. Introduzione all'etnomusicologia.* Torino: Einaudi.

Ruwet, N. (1972). Langage, musique, poésie. Paris: Seuil.

Sonic Visualizer. (2014). http://www.sonicvisualiser.org/

AUTOMATIC MUSIC TRANSCRIPTION OF THE *MAROVANY* ZITHER, BASED ON KNOWLEDGE FROM MUSICAL ACOUSTICS

Dorian Cazau, Olivier Adam
Sorbonne Universités / CNRS, UMR 7190
Institut Jean Le Rond d'Alembert, Equipe LAM
cazau@lam.jussieu.fr

Marc Chemillier
Ecole des Hautes Etudes en Science Sociale
Centres danalyse et de mathmatique sociales
chemilli@ehess.fr

1. EXTENDED ABSTRACT

Automatic Music Transcription (AMT) consists in automatically estimating the notes in a recording, through three attributes: onset time, duration and pitch. On the long range, AMT systems, with the purpose of retrieving meaningful information from complex audio, could be used in a variety of user scenarios such as searching and organizing music collections with barely any human labor (Klapuri, 2004). One common denominator of our different approaches to the task of AMT lays in the use of explicit music-related prior knowledge in our computational systems. A first step of this research project was then to develop tools to generate automatically this information. We chose not to restrict ourselves to a specific prior knowledge class, and rather explore the multi-modal characteristics of musical signals, including both timbre (i.e. modeling of the generic "morphological" features of the sound related to the physics of an instrument, e.g. intermodulation, sympathetic resonances, inharmonicity) and musicological (e.g. harmonic transition, playing dynamics, tempo and rhythm) classes. This prior knowledge can then be used in computational systems of transcriptions.

Our first axis of research requires a transcription accuracy high enough (i.e. average F-measure superior to 95 % with standard error tolerances) to provide analytical supports for musicological studies. Despite a large enthusiasm for AMT challenges, and several audio-to-MIDI converters available commercially, perfect polyphonic AMT systems are out of reach of today's algorithms. In this research project, we explore the use of multichannel capturing sensory systems for AMT of several acoustic plucked string instruments, including the following traditional African zithers: the *marovany* (Madagascar), the Mvet (Cameroun), the N'Goni (Mali). These systems use multiple string-dependent sensors to retrieve discriminatingly some physical features of their vibrations. For the AMT task, such a system has an obvious advantage in this application, as it allows breaking down a polyphonic musical signal into the sum of monophonic signals respective to each string (Cazau et al., 2013,?). Then, we come back to a monophonic transcription problem, which is considered as practically solved. For sake of flexibility and robustness, various sensor types (optical, piezoelectric and electromagnetic) have been comparatively tested. After experimentation, piezoelectric sensors, although quite invasive, prove to provide the best signal-to-noise ratio and multichannel separability. The development of this technology has allowed the constitution of a new sound dataset dedicated to AMT evaluation for plucked-string instrument repertoires. We gathered in these datasets audio recordings, MIDI-like transcripts and sound samples over the instrument pitch ranges. We also performed field recordings in Madagascar with local musicians during two missions (in July 2013 and 2014), using our multi-sensor retrieval systems.

Our second axis of research tackles the AMT task on audio recordings with more fundamental investigations on the use of prior knowledge in transcription performance, in regards to different plucked-string instrument repertoires. AMT is often divided into different processing stages, generally a multi-pitch estimation stage, followed by note segmentation and post-processing stages. In this research project, we mainly build our AMT framework from two methods, namely Probabilistic Latent Component Analysis (PLCA) (Smaragdis et al., 2006) for multi-pitch estimation and Hidden Markov Models (HMMs) for note segmentation and sequential post-processing. PLCA belongs to a spectrogram-factorization class of methods which is based on the modeling of a signal as a sum of basic elements. HMMs are a ubiquitous statistical tool to model time series data. We then develop different configurations of these methods to provide a powerful probabilistic framework covering the time-frequency domain on different time-scales, in which we develop original integration methods of prior knowledge. Timbre prior knowledge class is used to constrain generic signal models with acoustics-based information. A pitch-dependent sparsity prior has then been developed by modifying the EM update rules of PLCA, which is informed by the phenomenon of sympathetic resonances characterized acoustically. The use of pitch-wise multi-templates corresponding to different playing modes (e.g. dynamics, plucking techniques) have also been investigated, as well as the characterization of pitch activation profiles, which can be informed by specific temporal envelop modulations (e.g. intermodulation). For its part, musicological knowledge concerns more temporal musical structure and can be integrated conveniently into the HMM framework to build informed relations between frame-wise estimations. We develop original first- and second-order HMMs to model musicological polyphonic harmonic transitions between note mixtures, as well as higher-order HMMs including note duration modeling applied to note segmentation. This re-

search axis has first allowed achieving successful transcription enhancements in the *marovany* repertoires, by optimizing selectively the integration of musical knowledge.

Through our study of the *marovany* zither, we pave the way towards the development of AMT systems dedicated to other traditional plucked-string instruments. In future investigations, we will transpose this research framework to the repertoires of the Indian sitar and the Chinese GuQin zither.

Keywords : Automatic Music Transcription, Prior knowledge, Musical Acoustics, Non-eurogenetic music, Statistical learning, Computational Ethnomusicology

2. REFERENCES

Cazau, D., Adam, O., & Chemillier, M. (2013). An original optical-based retrieval system applied to music automatic transcription of the marovany cithara. In *Workshop on Folk Music Analysis, p. 44-50*.

Cazau, D., Chemillier, M., & Adam, O. (2013). Information retrieval of marovany zither music with an original optical-based system. In *Proceedings of DAFx 2013, Maynooth, Ireland*, (pp. 1–6).

Klapuri, A. (2004). Automatic music transcription as we know it today. *J. of New Music Research*, *33*, 269–282.

Smaragdis, P., Raj, B., & Shanshanka, M. (2006). A probabilistic latent variable model for acoustic modeling. In *Neural Information Proc. Systems Workshop, Whistler, BC, Canada*.

CONTRAST PATTERN MINING OF ETHIOPIAN BAGANA SONGS

Darrell Conklin[1,2] **Kerstin Neubarth**[3] **Stéphanie Weisser**[4]
[1] University of the Basque Country UPV/EHU, San Sebastian, Spain
[2] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
[3] Canterbury Christ Church University, United Kingdom
[4] Université Libre de Bruxelles, Brussels, Belgium
darrell.conklin@ehu.eus

1. INTRODUCTION

Contrast pattern mining (Dong & Bailey, 2012) aims to find patterns that contrast between groups in a dataset, even if those patterns cover just a small subgroup of instances. It is a relatively new area of data mining that is focused on *descriptive* rather than *predictive* methods, and on *interestingness* of discovered rules and patterns rather than *classification accuracy*.

In contrast pattern mining, a pattern is considered interesting if it differentiates between groups. In folk music analysis, groups can be given by e.g. geographical regions, folk music genres or tuning systems. Patterns can distinguish a target group (corpus) from other groups (anticorpus) in the dataset as positive patterns, which are *over*-represented in the corpus compared to the anticorpus (Conklin & Anagnostopoulou, 2011), or as antipatterns, rare patterns which are statistically *under*-represented in the corpus compared to the anticorpus (Conklin, 2013).

The *bagana* is a lyre played by the Amhara, a people settled mostly in Central and Northern Ethiopia. In a recent computational study of bagana songs Conklin & Weisser (2014) adapted antipattern discovery to find rare patterns when no explicit anticorpus is available. The method is able through a refinement search to explore the entire space of candidate patterns for minimal, statistically significant antipatterns. The method was capable of finding rare motifs reported to Weisser (2005) by the bagana master Alemu Aga. In that work, antipatterns are discovered which are under-represented in the *full corpus*. In the current work contrast pattern mining is applied to discover patterns significantly over-represented in events *played in a particular scale* or *played by a particular musician*.

2. DATA AND METHODS

The Ethiopian lyre *bagana* is a lyre with 10 strings, whose open strings are plucked with the left hand. A small dataset of 37 bagana songs has been encoded in terms of finger number sequences (isomorphic to a pitch representation). The dataset contains 1906 events, and represents seven different musicians and songs played in two different scales (called *tezeta* and *anchihoye*).

Following the notation of Neubarth & Conklin (2015), we consider boolean features G and X representing a *group* G and a *pattern* X. Referring to Table 1, the *support counts* of various combinations of X and G are: $n(X)$, the number of events initiating the pattern X; $n(G)$, the number of events in pieces within group G (i.e. in pieces played by a particular musician or played in a particular scale); $n(X \wedge G)$, the number of events initiating the pattern X in pieces within a group G, and so on. The total number of events in the dataset is given by N.

The goal of contrast pattern mining can be stated as: given a group G, find all patterns X that are significantly over- or under-represented in that group. An appropriate statistical test for this is Fisher's exact test on a 2×2 contingency table (using the shaded cells of Table 1). This test gives the probability of drawing $n(X)$ events from a total of N events, and finding $n(X \wedge G)$ events in the group G; the left or right tails of the distribution give the probability of an under- or over-represented pattern. Since there may be many significant patterns, those that are *minimal* (not containing another significant pattern) are discovered and presented. Minimal patterns will also have equal or higher support counts in the group than any more specific pattern.

3. RESULTS

The mining method was used to find over-represented patterns in the corpus in two separate configurations: pattern to musician; pattern to scale. Minimal patterns were mined using a significance level threshold of $\alpha = 0.001$. To avoid artefactual intra-opus repetition, patterns were required to occur in at least two different pieces.

Regarding personal style. It can be noted (Table 2, left) that certain patterns are primarily used by one of the seven musicians. The $[5]$ pattern and also significantly its immediate repetition $[5, 5]$ are over-represented in songs played by Akalu Yossef (77% of $[5, 5]$ events are played by Akalu Yossef). This characteristic can be linked to Akalu Yossef's specific background, as he is originally a *krar* player (another type of Amhara lyre) who has adapted elements of krar playing to bagana, including repetitive plucking of tones: such an extensive use of $[5, 5]$ might therefore be a marker of Akalu Yossef's personal style of playing. Regarding the over-representation of the patterns $[1, 3]$ and $[2, 4]$ in events played by Alemu Aga: he plays 54% of $[1, 3]$ and $[2, 4]$ events. Interestingly, both of those patterns were reported by Alemu Aga to Weisser (2005) as being frequent didactic motifs.

Table 1: Left: schema for a contingency table describing all relations between a pattern X and a group G. In the cells are the support counts for the different combinations of X and G. In light gray are the four parameters of Fisher's exact test. Right: the relations depicted spatially. The inner box contains the events initiated by the pattern X.

	G	$\neg G$	
X	$n(X \wedge G)$	$n(X \wedge \neg G)$	$n(X)$
$\neg X$	$n(\neg X \wedge G)$	$n(\neg X \wedge \neg G)$	$n(\neg X)$
	$n(G)$	$n(\neg G)$	N

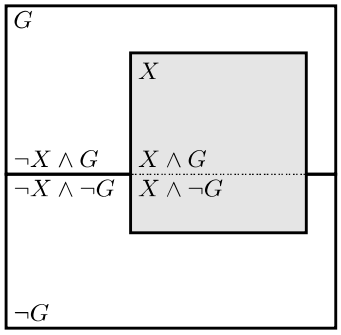


Table 2: Selected patterns appearing in at least two pieces whose usage is significantly ($\alpha = 0.001$) over-represented in the songs of the indicated musician (left) or scale (middle) when contrasted with the remaining musicians or other scale. Right: the scale used by songs of the indicated musician. †not minimal. $^{\star}n(X \wedge \neg G) = 0$

pattern	musician
$[5]$	Akalu Yossef
$[5,5]^{\dagger}$	Akalu Yossef
$[5,1]$	Abiy Seyoum
$[1,5]$	Abiy Seyoum
$[1,1]$	Alemu Aga
$[2,4]$	Alemu Aga
$[1,3]$	Alemu Aga
$[4,4]$	Yetemwork Mulat
$[3,1,3]$	Tafese Tesfaye
$[3,1,1,5]$	Yetemwork Mulat
$[2,3,1,1]$	Yetemwork Mulat

pattern	scale
$[5,1]$	anchihoye
$[5,5]^{\star}$	tezeta
$[1,5]$	anchihoye
$[3,2,4,2]^{\star}$	tezeta

musician	scale
Tafese Tesfaye	tezeta
Yetemwork Mulat	tezeta
Sosenna Gabreyesus	tezeta
Akalu Yossef	tezeta
Alemu Aga	tezeta
Ameha Desta	anchihoye
Abiy Seyoum	anchihoye

Regarding scale. It is interesting to note (Table 2, middle) that bigrams comprising the interval between 1 and 5 ($[5,1]$ and $[1,5]$) occur in anchihoye more frequently than expected. It could therefore be hypothesized that these specific patterns could be a marker for the sonorous identity of the scales. Such information would be extremely useful, as Amhara scales are complex to define in terms of interval sizes only: scholars have already noted that the latter is probably not the only defining element, especially for anchihoye, the more variable scale in terms of intervals (Weisser & Falceto, 2013).

4. DISCUSSION AND CONCLUSIONS

Analyses of associations between specific patterns and groups in the dataset (namely scales and musician) present a nuanced picture. As particular pieces in the corpus are played in only one scale, it is complex to determine with certainty if an over-representation of a pattern is due to a personal style of playing or to a musician's use of the scale. For example, the patterns [1,5] and [5,1] are over-represented in the playing of Abiy Seyoum, in the dataset represented only with pieces in anchihoye (Table 2, right): hence if these bigrams were markers of anchihoye, their occurrence in the musician's playing could be mediated (Neubarth et al., 2013) by the choice of scale.

It is important to state that no injunction nor interdiction is formulated by musicians in the bagana musical system (and the Amhara musical system in general). Also, realizations of songs vary a lot, without these differences being considered pertinent by players and listeners: according to context (audience's attentiveness, musician's tiredness, condition of the instrument, etc.), bagana players will adjust their performance. Moreover, several bagana melodies are transmitted from master to pupil (entirely new melodies are relatively rare). Most musicians adopt, out of respect, their master's playing style. They however "personalize" the songs, by adding or removing patterns or single notes, hence modifying pattern frequencies.

The flexibility and complexity of the bagana musical system requires therefore a complementary approach to statistical analysis. In a next step, integration of the musical context of the patterns could provide interesting information: a pattern at the beginning or the end of a musical sequence might be of a different significance than a pattern in the middle part, for example. A field experimentation with musicians and generated bagana song models could provide more information regarding the pertinence of these results.

5. ACKNOWLEDGEMENTS

This research is supported by the project Lrn2Cre8 which is funded by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 610859.

6. REFERENCES

Conklin, D. (2013). Antipattern discovery in folk tunes. *Journal of New Music Research*, *42*(2), 161–169.

Conklin, D., & Anagnostopoulou, C. (2011). Comparative pattern analysis of Cretan folk songs. *Journal of New Music Research*, *40*(2), 119-125.

Conklin, D., & Weisser, S. (2014). Antipattern discovery in Ethiopian bagana songs. In S. Dzeroski, P. Panov, D. Kocev, & L. Todorovski (Eds.), *17th International Conference on Discovery Science* (Vol. 8777, pp. 62–72). Bled, Slovenia: Springer.

Dong, G., & Bailey, J. (Eds.). (2012). *Contrast Data Mining: Concepts, Algorithms, and Applications*. Chapman and Hall/CRC.

Neubarth, K., & Conklin, D. (2015). Contrast pattern mining in folk music analysis. In D. Meredith (Ed.), *Computational Music Analysis.* Springer. (To appear)

Neubarth, K., Johnson, C. G., & Conklin, D. (2013). Discovery of mediating association rules for folk music analysis. In *5th International Workshop on Music and Machine Learning at ECML/PKDD 2013 (MML 2013).* Prague, Czech Republic. (w/o pages)

Weisser, S. (2005). *Etude ethnomusicologique du bagana, lyre d'ethiopie* (Unpublished doctoral dissertation). Université Libre de Bruxelles.

Weisser, S., & Falceto, F. (2013). Investigating qeñet in Amhara Secular Music: An Acoustic and Historical Study. *Annales d'Ethiopie 28*, 299-322.

COMPUTATIONAL TEXTSETTING ANALYSIS OF DUTCH FOLK SONGS

Varuṇ deCastro-Arrazola[1,2], **Peter van Kranenburg**[1], **Berit Janssen**[1,3]
[1]Meertens Instituut, Amsterdam [2]Universiteit Leiden [3]Universiteit van Amsterdam
{varun.decastro-arrazola,peter.van.kranenburg,berit.janssen}@meertens.knaw.nl

ABSTRACT

The study of textsetting describes how words and tunes are aligned in songs. In languages with word stress such as English, it has been shown that the metrical prominence of melodies is aligned in a non-random way with the lyrics' word stress (Dell & Halle 2009). The present study addresses the textsetting rules of Dutch using a large dataset of folk songs and a novel methodology. The main findings are the following: (1) the combination of linguistic stress and metrical prominence moving in opposite directions is avoided; (2) textsetting rules do not apply across phrases, but they do across words; (3) the avoidance of stress and prominence combinations depends on factors such as phrase-finality and the presence of melisma. In contrast with previous textsetting studies, our approach does not use pre-defined mismatches, but induces the avoided stress / prominence combinations in a data-driven way. This allows for a more systematic understanding of how words and tunes are to be aligned in a given tradition.

1. INTRODUCTION

Songs can be perceived globally as homogeneous objects, but we can also consider them composite objects with two main components: a text and a tune. The analysis of textsetting describes how these two components are combined. Textsetting rules state whether particular combinations of linguistic and musical features are preferred or avoided in a given musical tradition.

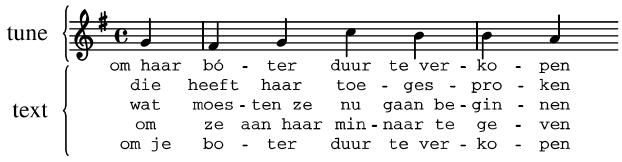


Figure 1: Songs can be analysed as composite objects combining a text and a tune. This tune (with slight variations) is used several times in the same song, while different words are set to it.

The autonomous status of the tune can be observed, for instance, in cases where the same melody is used several times with varying lyrics. Figure 1 shows such an example taken from the MTC-FS corpus (van Kranenburg et al. 2014). Tunes can therefore be considered abstract templates to which words are set.

Avoiding stressed syllables in weak musical positions, for instance, is a common tendency in languages such as English or French (Dell & Halle 2009). In tone languages, where pitch contours are used to distinguish lexical or grammatical meaning (e.g. Chinese), textsetting rules can specify how to match linguistic and musical pitch contours (Schellenberg 2013).

Dutch is a non-tonal language, but it does have word stress; e.g. the first syllable in a word like *boter* 'butter' is more prominent than the second syllable. Similarly, in metered music some positions are more prominent than others. We will use *stress* to talk about linguistic prominence, and *prominence* to talk about musical prominence.

The goal of this paper is twofold. From a methodological point of view, it presents a systematic way of addressing the textsetting problem computationally (virtually all existing studies on the topic rely on manual analyses; cf. Temperley & Temperley 2013). Secondly, it provides a first description of textsetting in Dutch folk songs.

2. METHOD

2.1 Material

In order to study the textsetting rules of Dutch folk songs we analysed 3724 songs from the MTC-FS corpus (van Kranenburg et al. 2014). Most of the songs were collected through fieldwork between the 1950s and the 1990s as part of the radio programme *Onder de groene linde* led by Ate Doornbosch. The corpus also contains similar songs taken from 19th and 20th century songbooks. [1]

The original corpus contains 3861 songs. However, the features we focus on (stress and prominence) were not always obtainable. Songs encoded as having free rhythm (n = 125) were excluded because they lack the feature of prominence. Linguistic stress for the lyrics was obtained through a nearest-neighbour lookup in the e-Lex [2] lexical database (as specified in van Kranenburg & Karsdorp 2014). Thus, the database lookup is robust against minor variations in spelling. Cases in which the nearest neighbour in the e-Lex database has a different number of syllables than the word in the lyrics were discarded. Any phrase containing one such word has also been excluded from the analysis (n = 2451 phrases).

Every song in the dataset is divided into stanzas; stanzas are divided into phrases; phrases contain notes, which can then be associated to syllables. For the purposes of this pa-

[1] The melody, text and metadata for each song is openly available in several formats at www.liederenbank.nl/mtc.

[2] http://tst-centrale.org/en/producten/lexica/e-lex/7-25

per, stanzas are roughly equivalent to strophes, and phrases are also referred to as lines. The filtered dataset contains 3724 songs, 3973 stanzas, 20662 phrases, 185263 notes, and 176708 syllables. Syllables and notes are often in a one-to-one correspondence. Some syllables, though, span over more than one note; such syllables are referred to as *melismas*. In the filtered dataset, 4.46% of the syllables are melismas.

2.2 Corpus annotation

Stress is not a feature present in the original dataset, it was looked up at the e-Lex database. Stress is encoded in a binary way in the database; each syllable gets a value of either `0` (unstressed) or `1` (stressed). Secondary stress is not encoded. Figure 2 illustrates how this and the following features related to stress and prominence have been automatically annotated.

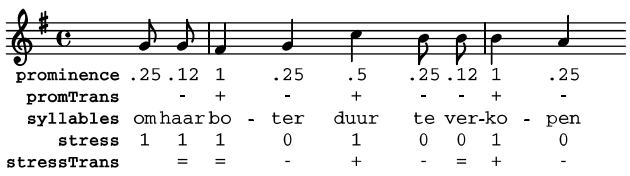


Figure 2: Sample annotation of stress, prominence and their respective contours.

Musical prominence is also not explicitly encoded as a feature in the MTC-FS dataset. However, this feature can be inferred from the symbolic representation of the tunes. For each note, we know its position within the musical bar, and the time signature this bar belongs to (e.g. 6/8). Given that information, relative prominence can be derived. This was done using the `music21` software (Cuthbert & Ariza 2010). Prominence values range from `0` to `1`, the first position of the bar being assigned a `1`.

Both stress and prominence are relative notions, that is, given a syllable in isolation, its raw stress/prominence value is trivial. Hence, to capture how stress and prominence are aligned, it becomes necessary to compare a syllable with its neighbours. We have achieved this by computing the transition for the stress and prominence values of each syllable compared to the values of the preceding syllable. This produces three possible stress/prominence contours: decreasing (–), same (=) and increasing (+), as illustrated in Figure 2. Note that the first syllable of a song does not have any preceding syllable to compare its stress or prominence to; that is, its transitions values are left empty (set to `NA`).

2.3 Statistical analyses

In order to find out which combinations of stress and prominence are being avoided in the corpus, we first computed the frequency of each of the 9 possible transition combinations (decrease, same, increase for each feature). We then took the marginal probabilities of the contingency table to calculate the expected frequencies for each cell. Last, we calculated the ratio between observed and expected frequencies in order to detect under-represented combinations.

For instance, the proportion of syllables having a decreasing stress and an increasing prominence is 0.03. If we look at the marginal frequencies, we observe that 0.26 of all syllables show a decreasing stress, and 0.46 show an increasing prominence. The expected frequency for the combination of the two is the product of the marginal frequencies: $0.26 \times 0.46 = 0.12$. We can now compare the observed frequency (0.03) with the expected frequency (0.12), and note that that particular combination of features is under-represented in the dataset; i.e. if the alignment of text and tune was done randomly, that combination would be more frequent.

As a way of assessing the degree to which a combination is over- or under-represented, we calculated an association factor by dividing the observed frequency by the expected frequency. In the above example, this yields 0.26. Association factors can range from 0 to infinity, where 1 indicates that the combination is not controlled, as it occurs at chance level. We take the conventional threshold of 0.5 as a cut-off point to select combinations which are significantly avoided (Agresti 2013).

The same kind of analysis was performed on conditioned subsets of the data in order to refine the textsetting rules in two ways. First, the domain on which textsetting rules operate was addressed. Given that our analyses are computed on two-syllable windows (by comparing every syllable with its preceding syllable), some of these windows will go across boundaries. Figure 3 illustrates the two cross-boundaries contexts which were analysed: beginning of phrases and beginning of words. Second, we investigated whether the alignment of stress and prominence is influenced by other features, namely, being sung to a melisma, and being placed at the end of phrases.

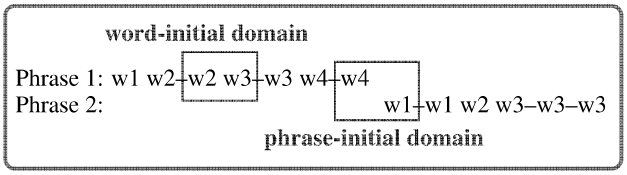


Figure 3: We define the textsetting domain with a two-syllable window. In the above example, the word-initial domain includes the last syllable of word 2 (w2) and the first syllable of word 3 (w3). The phrase-initial domain includes the last syllable of phrase 1 and the first syllable of phrase 2.

3. RESULTS

Table 1 shows the association factors for all possible combinations of stress and prominence transitions. Two of the nine possible alignments are being avoided: [stress+, prom–], and [stress–, prom+]. That is, when aligning words to a tune, stress and prominence should not go in opposite

directions. When stress or prominence show a = contour, association values are close to 1, suggesting these alignments are not controlled.

	stress-	stress=	stress+
prom-	1.77	0.93	**0.36**
prom=	0.85	0.95	1.24
prom+	**0.26**	1.08	1.60

Table 1: Association factors for stress and prominence contours. Combinations below the 0.5 threshold, marked in bold red, are considered avoided alignments.

3.1 Relevant domain for textsetting

Next, we use this measure of association to further investigate the domain at which textsetting rules operate. Is the alignment of stress and prominence controlled across phrases? In order to address this question, we divided the data into phrase-initial contexts (illustrated in Figure 3) and non-initial contexts, and then calculated the association factors between stress and prominence.

The right panel in Figure 4(a) shows that the transition of prominence from the last syllable of a phrase to the first syllable of the following phrase is not associated to the stress transition of those two syllables (values do not go below the 0.5 threshold). Alignment of stress and prominence is controlled within phrases (left panel), but not across them (right panel).

In what follows, transition values for phrase-initial contexts are discarded, because they do not appear to be actively controlled. In the left panel in Figure 4(a), we also observe that both stress– and stress+ mismatches are now equally avoided within phrases, which was not the case in Table 1, where phrase-initial contexts were included in the analysis.

The second question on the textsetting-relevant domain asks whether stress and prominence transitions are controlled across words. Figure 4(b) shows that alignment is controlled both within and across words. Besides, there is one additional constraint when aligning stress and prominence word-internally (left panel): [stress+, prom=] is avoided. That is, in polysyllabic words, an increase of stress must be aligned with an increase of prominence.

3.2 Additional textsetting factors

How do other factors make stress / prominence mismatches more or less acceptable? Figure 6(a) shows the data divided into two subsets: phrase-final and non-final contexts. Mismatches involving stress+ are much more avoided phrase-finally (right panel), while the ones with stress– are more accepted in this context. Figure 5 displays one of the rare examples of a [stress+, prom–] at the end of the phrase.

In Figure 6(b) the data has been divided according to whether the current syllable has or has not a melisma (m1, m0), and whether the preceding syllable has or has not a melisma (mp1, mp0). The condition where a syllable has

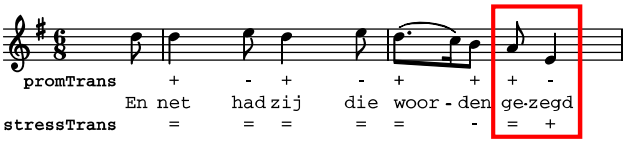


Figure 5: The alignment [stress+, prom–] is much more avoided phrase-finally than elsewhere.

a melisma *and* is preceded by a melisma (m1mp1) is excluded from the plot because it turns out to be statistically uninformative due to its infrequency.

When a syllable is sung to a melisma (rightmost panel), [stress+, prom–] mismatches are even more infrequent than without a melisma. In contrast, [stress–, prom+] mismatches become more acceptable when sung to a melisma (while still being avoided).

Regarding the effect of a preceding melisma, the most salient one is shown in the left column of the middle panel of Figure 6(b). A [stress–, prom+] alignment is not avoided if the first of the two syllables is a melisma. This is illustrated in the penultimate word of Figure 5 (*woorden*).

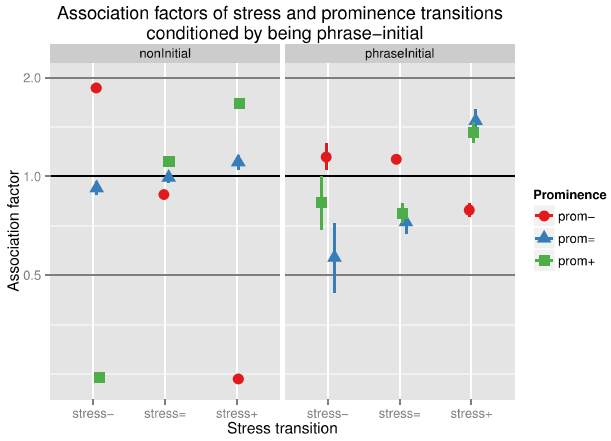
4. DISCUSSION

The avoided combinations highlighted in the previous section can be reformulated in terms of constraints which are active when a Dutch speaker creates a new song, and inversely when a song is judged as well- or ill-formed. Minimally, the window where stress and prominence are compared must comprise two adjacent syllables. This window is reset afresh at the beginning of each phrase, but not at the beginning of words. Further analysis should determine whether the relevant textsetting domain is smaller than the musical phrase, maybe corresponding to a phonological phrase (Proto & Dell 2013).

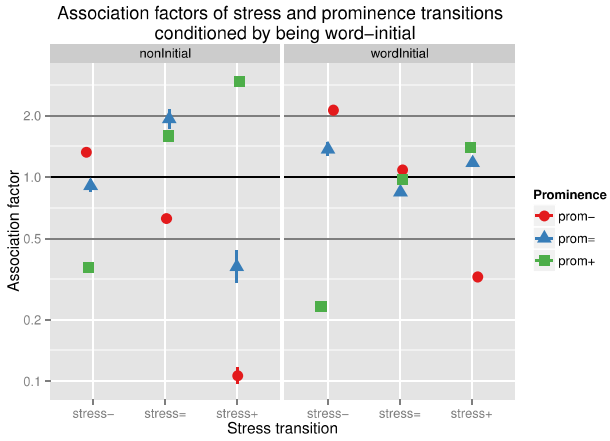
Within this two-syllable domain, the general rule is that stress and prominence should not move in opposite directions. The [stress–, prom+] misalignment constraint can be relaxed phrase-finally and when the second syllable is sung to a melisma. If the first syllable is sung to a melisma, the constraint does not apply at all. The [stress+, prom–] misalignment constraint is *stricter* phrase-finally, and when the second syllable is sung to a melisma.

The alignment of stress+ with prom= is also avoided within polysyllabic words, a constraint not explicitly mentioned in the previous textsetting literature. This might suggest that two adjacent positions with the same degree of prominence according to a traditional analysis (e.g. positions 2 and 3 of a 6/8) are actually not treated as equal, but rather the second one being weaker, hence asking for a less stressed syllable.

Future work will determine whether Dutch listeners are sensitive to the rules stated above. Besides, the reason why constraints are relaxed in certain contexts remains open. For instance, the expectation that phrases should finish in a more prominent note (because of closure) may account for the observed phrase-finality effects.

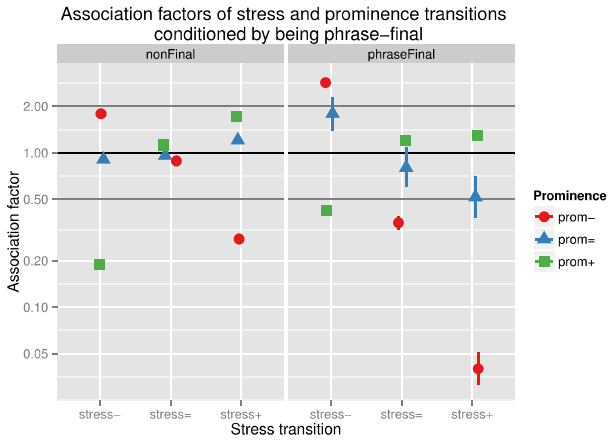


(a) Association between stress and prominence contours for syllables in phrase-initial and non-initial contexts.

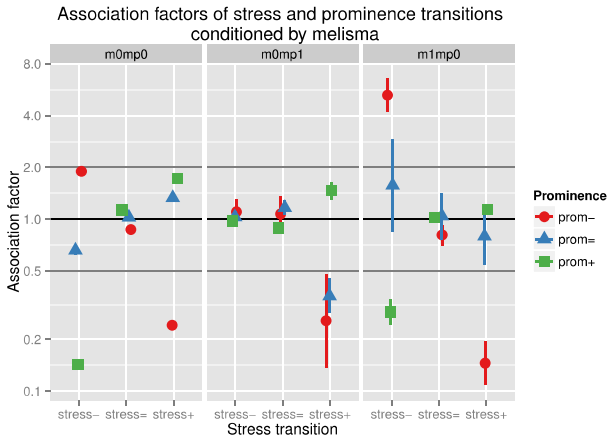


(b) Association between stress and prominence contours for in word-initial and non-initial contexts.

Figure 4: Association factors for boundary-crossing contexts compared to non-crossing contexts: (a) phrase-initial context, (b) word-initial context. A value of 1 indicates random alignment. Upper and lower significance thresholds are set at 2 and 0.5.



(a) Association between stress and prominence contours for syllables in phrase-final and non-final contexts.



(b) Association between stress and prominence contours for syllables with(out) melisma (m1, m0), preceded by syllables with(out) melisma (mp1, mp0).

Figure 6: Effect of (a) phrase-finality and (b) melisma on the association between stress and prominence contours. A value of 1 indicates random alignment. Upper and lower significance thresholds are set at 2 and 0.5.

In order to follow up the present study, further linguistic and musical features can be considered. Dutch is said to have secondary stress, which means that some words contain more than one stressed syllable (Booij 1995). This could account for some of the examples currently analysed as mismatches. On the musical side, note duration and pitch can be of particular interest, since they can contribute to the perception of prominence (Müllensiefen et al. 2009). Also, the effect of factors such as melisma or phrase-finality have been studied here in isolation; however, more complex interactions between them and with other features can be expected.

A detailed description of this kind of interactions can eventually shed some light on the cognitive processes involved when simultaneously processing music and language. The fact that we unconsciously control how linguistic prosody and musical features are to be aligned may suggest that these two domains are processed by the same neural resources (Zatorre & Baum 2012; Hausen et al. 2013).

Finally, an understanding of how words and tunes are aligned in a given language provides the basis for a number of applications. For instance, historical lyrics for which the tune is unknown can be matched to well-aligned melodies of the same period (and vice versa for text-less tunes). In order to address this kind of automatic alignment, the observed avoidance values can be straightforwardly converted to weighted penalties.

5. CONCLUSION

This paper constitutes an initial study on the textsetting of Dutch folk song. We show that, as in other languages, linguistic stress and musical prominence moving in opposite directions is avoided. The domain where this constraint is active is smaller than the musical phrase, and bigger than the word. A number of factors (phrase-finality, melisma) make these mismatches more or less acceptable. The proposed methodology has the advantage of being able to infer avoided combinations of linguistic and musical features, and to make explicit predictions about which contexts would be perceived as more or less ill-formed. Careful inspection of the existing songs where avoided combinations occur remains crucial in order to systematically test further features which may interact with stress and prominence.

6. REFERENCES

Agresti, A. (2013). *Categorical Data Analysis*. Hoboken (New Jersey): John Wiley & Sons.

Booij, G. (1995). *The Phonology of Dutch*. Oxford: Clarendon Press.

Cuthbert, M. S. & Ariza, C. (2010). music21: A toolkit for computer-aided musicology and symbolic music data. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)* (pp. 637–642). ISMIR.

Dell, F. & Halle, J. (2009). Comparing musical textsetting in french and english songs. In J.-L. Aroui & A. Arleo (Eds.), *Towards a Typology of Poetic Forms* (pp. 63–78). Benjamins.

Hausen, M., Torppa, R., Salmela, V. R., Vainio, M., & Särkämö, T. (2013). Music and speech prosody: a common rhythm. *Frontiers in psychology*, *4*, 1–16.

Müllensiefen, D., Pfleiderer, M., & Frieler, K. (2009). The perception of accents in pop music melodies. *Journal of New Music Research*, *38*(1), 19–44.

Proto, T. & Dell, F. (2013). The structure of metrical patterns in tunes and in literary verse. evidence from discrepancies between musical and linguistic rhythm in italian songs. *Probus*, *25*(1), 105–138.

Schellenberg, M.-H. (2013). *The Realization of Tone in Singing in Cantonese and Mandarin*. PhD thesis, The University of British Columbia.

Temperley, N. & Temperley, D. (2013). Stress-meter alignment in french vocal music. *Journal of the Acoustical Society of America*, *134*(1), 520–527.

van Kranenburg, P., de Bruin, M., de Grijp, L., & Wiering, F. (2014). The meertens tune collections. *Meertens Online Reports*, *1*.

van Kranenburg, P. & Karsdorp, F. (2014). Cadence detection in western traditional stanzaic songs using melodic and textual features. In *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)* (pp. 391–396). ISMIR.

Zatorre, R. J. & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune. *PLoS Biology*, *10*(7), e1001372.

EFFICIENT ALGORITHMS FOR MELODIC SIMILARITY IN FLAMENCO SINGING

J.M. Díaz-Báñez
Universidad de Sevilla, Spain
dbanez@us.es

N. Kroher
Univ. Pompeu Fabra, Spain
nadine.kroher@upf.edu

J.C. Rizo
Universidad de Sevilla, Spain
juarizmas@us.es

ABSTRACT

Alignment algorithms are widely used as a measure to compare similarities of symbol sequences and have previously been employed in the context of melodic similarity among pitch contours. In the particular case of flamenco singing, each style is characterised by a common melodic skeleton and melodic sequences among excerpts belonging to the same style should thus yield high similarity values. However, the absence of scores and the spontaneous improvisational character of the style (ornamentation, prolongation, melodic variation) add high complexity to the problem. In the present study, we employ Dynamic Time Warping (DTW) to perform pair-wise alignment of melodic contours and estimate their similarity from the resulting alignment cost. In order to reduce the time and space complexity of the alignment procedure and to eliminate microtonal ornamentations as a source of error, we aim to pre-process pitch contours by means of an efficient and musically meaningful segmentation. We consider a particular type of flamenco, the *tonás*, a sub-genre of a cappella songs performed in free rhythm and characterised by high degree of complex ornamentation. We assess the quality of the obtained similarity measures for a variety of pre-processing strategies by analysing the intra-cluster distances and visualise the results using phylogenetic trees. Our study shows, that a meaningful contour pre-processing by means of note-level segmentation can significantly reduce the computational complexity and increase the overall performance.

1. INTRODUCTION

Characterising the melodic similarity among audio recordings is a major challenge in the music information retrieval community and sets the basis for a number of related tasks, such as automatic style classification or melodic pattern detection. In the context of flamenco singing, the absence of scores and the presence of ornamentations and vibrato adds complexity to the problem. We focus on the particular case of the *tonás*, an a cappella flamenco singing style: A set of distinct melodic skeletons are performed in free rhythm and are subject to melodic variation and a high degree of complex micro-tonal ornamentation. Given these characteristics, this style represents a particular challenge for the automatic characterisation of melodic similarity, which has previously been studied in the scope of the COFLA project [1]. A first approach was presented in Cabrera et al. (2008): Both manual and automatic transcriptions are used to generate a simplified symbolic representation of the melodic content of a selection of 40 a cappella singing recordings. Pair-wise melodic similarity is measured based on distances among note interval vectors. Furthermore, phylogenetic trees computed from the resulting distance matrix are analysed regarding style organisation and historic evolution. A more detailed analysis is provided in Mora et al. (2010), where the distance calculations are based on expert-defined high-level features. Exploring perceptual aspects of melodic similarity (Kroher et al., 2014), human similarity judgements among a cappella recordings are modelled by analysing statistical performance descriptors extracted from automatic transcriptions.

All of the mentioned previous approaches rely either on manual annotations and consequently intense human effort or on a state of the art automatic transcription system (Gómez & Bonada, 2013), which is associated with a certain computational cost and low accuracy regarding transcription performance. In this paper, we present a melodic similarity measure based on fundamental frequency contours: Given the melodic contours of two excerpts, we use *dynamic time warping* to determine the best possible alignment and use the resulting cost as a measure for the melodic similarity among them. This technique is suitable for this task given that structural alignment is a prominent model in cognitive science for human perception of similarity (Goldstone & Son, 2005) and it has previously given good results when applied in the context of melodic similarity (Mongeau & Sankoff (1990); Pikrakis et al. (2003); Molina et al. (2013); Díaz-Báñez & Rizo (2014)). Based on the obtained distance measures, we can visualise and interpret intra-style clusters using phylogenetic trees. In a comparative study, we investigate the influence of different contour pre-processing algorithms by means of computational cost and classification accuracy and observe that an appropriate simplification of the contour yields to an improvement regarding both aspects.

2. METHODOLOGY

Our goal is to develop an efficient and effective algorithm for melodic similarity among a cappella flamenco songs based on temporal alignment using *dynamic time warping* (DTW). Instead of applying DTW to the raw pitch contour, we first obtain an approximation of the signal by means of converting it into piece-wise constant segments. We propose several methods and their combinations for contour pre-processing and evaluate regarding computational efficiency and classification accuracy. The experimental setup is depicted in Figure 1 and the involved signal processing stages are described below.

[1] http://mtg.upf.edu/research/projects/cofla

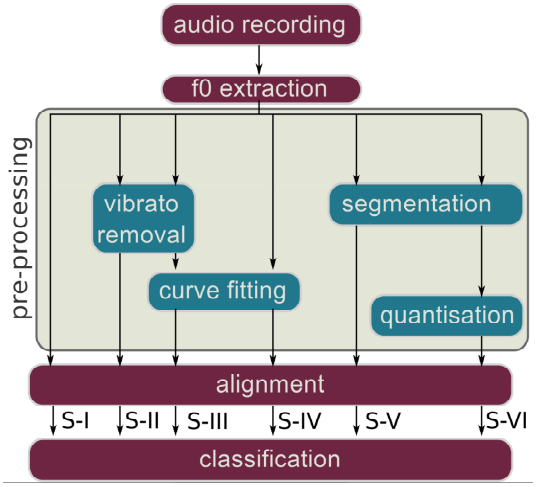


Figure 1: Experimental setup.

2.1 Fundamental frequency extraction

We use a state of the art predominant melody extraction algorithm described in Salamon & Gómez (2011) to estimate the fundamental frequency corresponding to the vocal melody. In order consider only voiced segments, we remove spurious notes and frames in which no predominant contour was detected.

2.2 Contour pre-processing

We explore several pre-processing stages and various combinations among them. The involved processing blocks are explained below:

2.2.1 *Vibrato removal*

Flamenco singing is characterized by en extensive use of vibrato, not only on steady long notes, but also during rapid melodic movements and melisma. The singing voice vibrato can be modeled as a periodic fluctuation of the pitch contour, which can extend more than one semitone. Such fast pitch fluctuations aggravate the process of aligning pitch contours and consequently of characterizing melodic similarity. We therefore investigate, in how far detecting and eliminating vibrato from the fundamental frequency envelope can benefit the dynamic temporal alignment by means of computational cost as well as precision.

In order to detect vocal vibrato in the pitch contour, we apply a procedure similar to the method described in Herrera & Bonada (1998) with some additional constrains: We analyse the pitch contour $f[n]$ in frames of $34ms$ with a temporal overlap of 95%. The fast fluctuations are isolated by subtracting the mean pitch from the contour segment. Subsequently, we compute the autocorrelation R_{yy} over the frame of length N and search for local maxima between a lag l of 140 and $156ms$, in order to cover a vibrato range from 4 to $9Hz$:

$$R_{yy}[l] = \sum_{n=n_{start}}^{N-h-1} f[n] \cdot f[n+l] \qquad (1)$$

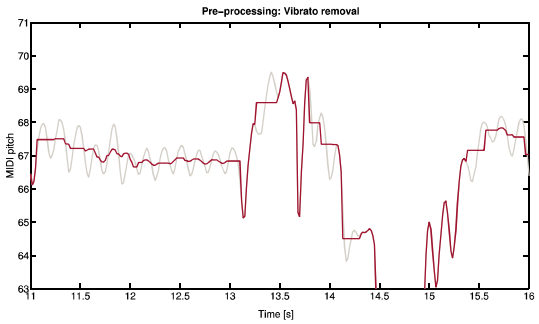


Figure 2: f0 contour before (light) and after (strong) vibrato reduction.

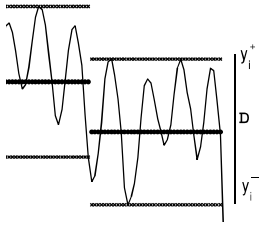


Figure 3: The fitting algorithm: f0 (fine line), segmented contour (strong line), upper and lower bounds (dashed line).

If a peak is found in this range, we can assume a periodic fluctuation with a period spanning from the start of the analysis frame n_{start} to the peak location n_{peak}. In order to eliminate vibrato without losing the underlying pitch information, we furthermore need to verify that no note change occurs within the vibrato period. We therefore apply a boundary condition based on the maximum pitch distance d_p in $cents$ within the estimated vibrato period:

$$d_p = 1200 \cdot log_2 \left(\frac{\max_{n_{start}<n<npeak} f[n]}{\min_{n_{start}<n<npeak} f[n]} \right); \qquad (2)$$

We assume that no note change takes place during the estimated vibrato period if $d_p < 200$ and eliminate the vibrato period be averaging the signal in the considered range.

2.2.2 *Curve fitting*

The contour is segmented into a set of constant segments where the absolute error between the constant and the contour does not exceed a given bound α. The algorithm, based on the approach of Díaz-Báñez & Mesa (2001), works as follows: Given a set of points $P = \{p_1, p_2, \cdots, p_n\}$ in the plane and an error tolerance α, plot vertical segments V_i of length $D = 2\alpha$ centered at each point p_i, refer to Figure 3. Note that the constraint that each point be within α of the step function is equivalent to saying that the step function intersects each of these segments of length $D = 2\alpha$. Thus, sweeping from left to right, the algorithm greedily tries to intersect as many consecutive segments as possible, before starting a new step and repeating this pro-

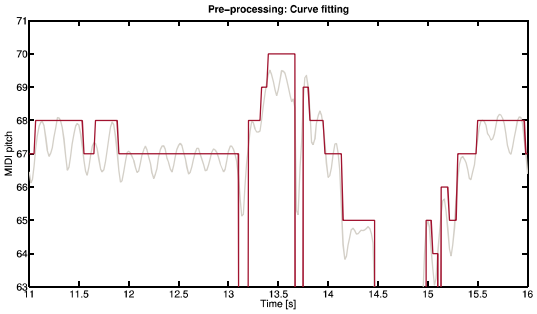


Figure 4: Contour segmentation with the fitting algorithm: f0 (fine line) and segmented contour (strong line).

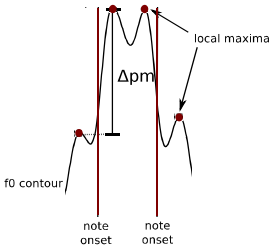


Figure 5: Pitch contour segmentation based on local maxima.

cedure. A vertical segment V_i defines a y interval $[y_i^-, y_i^+]$ where y_i^- and y_i^+ denote the y coordinates of the lower and upper endpoints, respectively. Sweeping from left to right, we maintain the intersection Δ of the y intervals of the vertical segments until we reach a segment V_i whose y interval does not intersect Δ, in which case we terminate the current step, and start a new step at V_i setting $\Delta = [y_i^-, y_i^+]$. Clearly, this algorithm is easy to implement, runs in $\Theta(n)$ time, and it constructs a step function with minimum number of horizontal segments for the error tolerance α.

2.2.3 Note-level segmentation

We furthermore propose an alternative simple signal processing method to estimate note boundaries directly from the unprocessed pitch contour. We assume that the underlying slowly varying pitch can be estimated from the upper envelope of the fundamental frequency trajectory. We propose a simple onset detection function $d[t]$ based on the relative pitch difference of adjacent local maxima, which has shown to be robust towards several common types of ornamentations. First, the extracted pitch contour is mapped to a cent scale with respect to a reference frequency $f_{Ref} = 55Hz$:

$$c[n] = 1200 \cdot log_2 \frac{f[n]}{f_{Ref}[n]} \tag{3}$$

Next, the relative pitch distance $\Delta pm_{i,i+1}$ among subsequent local maxima pm_i is calculated. If this distance exceeds 80 cents, we assume the beginning of a new note between both local maxima. The process is depicted in Figure 5. Within each estimated note, we quantise the pitch to its median frequency value.

2.2.4 Semi-tone quantization

Symbolic note representations, such as the MIDI standard, use pitch values quantised to the equal tempered semitone scale. Due to the pitch-continuous nature of the singing voice, the presence of ornamentations and vibrato, intonation errors and possible deviations from the standard tuning of $A_4 = 440Hz$, note quantization to semitones poses a challenge for the given material. Nevertheless, we want to explore if the temporal contour alignment ad consequently the style classification can be improved if contour segments are quantized to semitones. We first analyze each track and estimate a global tuning frequency using circular statistics, similar to the approach presented in Dressler & Strech (2007): For each cent contour frame $c[n]$, the local deviation $\Delta_{tune}[n]$ from the standard tuning A_4 can be calculated as

$$\Delta_{tune}[n] = c[n] - round\left(\frac{c0[n]}{100}\right) \cdot 100. \tag{4}$$

We model this local deviation as an angle $\alpha[n]$ in the complex plane, circulating around $\pm 50 cents$:

$$\alpha[n] = \Delta_{tune}[n] \cdot \frac{2\pi}{100} \tag{5}$$

The averaging process now corresponds to a complex pointer summation over all N frames and the average tuning deviation can be estimated from the resulting angle.

$$\overline{\Delta_{tune}} = \arg\left(\sum_{n=1}^{N} cos(\alpha[n]) + j \cdot \sum_{n=1}^{N} cos(\alpha[n])\right) \tag{6}$$

The estimated global tuning frequency $A_{4,T}$ results to:

$$A_{4,T} = 2^{\frac{\overline{\Delta_{tune}}}{1200}} \cdot A_4 \tag{7}$$

Based on this estimation, we can compute a pitch histogram for each note segment, summing the occurrences of each semitone. First, the cent contour $c[n]$ is re-estimated with respect to the estimated tuning frequency, similar to Eq. 3. The centre values $C[k_s]$ of the semitone bins now correspond to multiples of 100 cents, e.g. the k^{th} semitone above and the $-k^th$ semitone below A4,T is given as:

$$C[k_s] = k_s \cdot 100. \tag{8}$$

We can now compute the semitone histogram $H[k_s]$ by quantizing each cent contour frame to its closest centre value. The contour is quantized within the note boundary to the semitone corresponding bin with the highest occurrence. From the maximum pitch histogram bin K, we can compute the corresponding quantized frequency value f_Q:

$$f_Q = 10^{\frac{C[K]}{1200 \cdot 3.322}} \cdot A_4 \tag{9}$$

Figure 6 shows the difference between a quantized and an un-quantized segmentation.

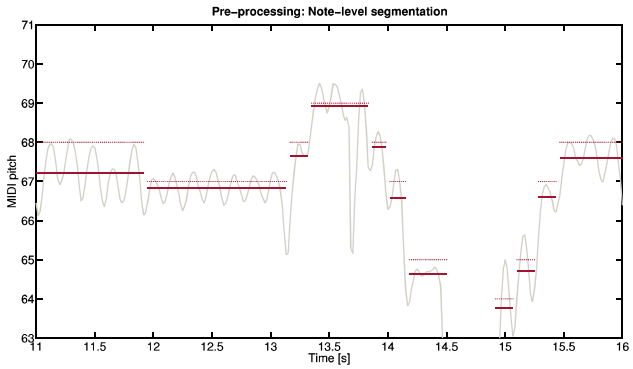


Figure 6: Pitch contour segmentation: f0 (light line), unquantized (strong line) and quantized (strong dashed line) melodic contour.

2.3 Alignment

We use *dynamic time warping* (Sakoe (1978); Myers et al. (1980)) to perform pair-wise alignment of pitch sequences. This technique has previously been used in the context of singing voice assessment Molina et al. (2013) to characterise similarity among pitch contours. The cost $m[i, j]$ of aligning the i^{th} sample of a cent-scaled contour $c_1[n]$ of length I to the j^{th} cent-scaled contour $c_2[n]$ of length J is given as:

$$m[i, j] = \sqrt{(c_1[i] - c_2[j])^2} \tag{10}$$

In the same manner, we can compute the cost matrix $M(I, J)$ containing each possible alignment combination. Using a dynamic programming algorithm, the best alignment path l of length L across the cost matrix M (refer to Figure 7) can be found by minimizing the global warp cost (Myers et al. (1980)):

$$m_p = \sum_{l=1}^{L} M[l_i, l_j] \tag{11}$$

The global warp cost furthermore serves as a distance measure to characterize the similarity among two sequences: The smaller m_p, the more similar the two sequences. The computational cost of finding the best path depends on the size of the M. A prior segmentation stage, i.e. curve fitting or note-level segmentation, thus results in a significant reduction of the dimension of M and consequently in a faster alignment procedure.

3. EVALUATION AND RESULTS

3.0.1 Database

Songs belonging to the *tonás* style can be further categorized in various sub-styles, each defined by a common underlying melodic skeleton. Consequently, song belonging to the same sub-style should lead to high melodic similarity values. We use a set of 24 *tonás* recordings taken from a publicly available dataset [2], containing twelve excerpts of the *martinete* and twelve excerpts of the *debla* sub-style. We refer to Mora et al. (2010) for a comprehensive description of the considered styles and their musical characteristics.

[2] http://mtg.upf.edu/download/datasets/tonas

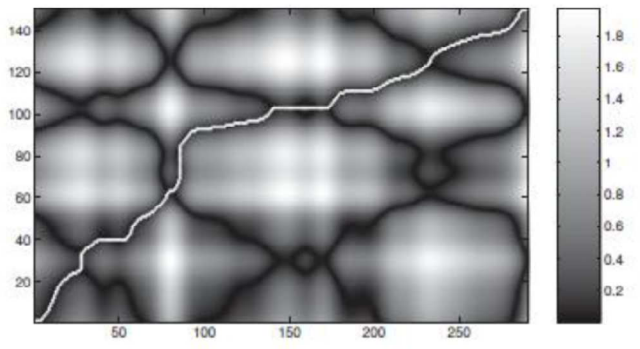


Figure 7: Cost matrix M and optimum alignment path l (white).

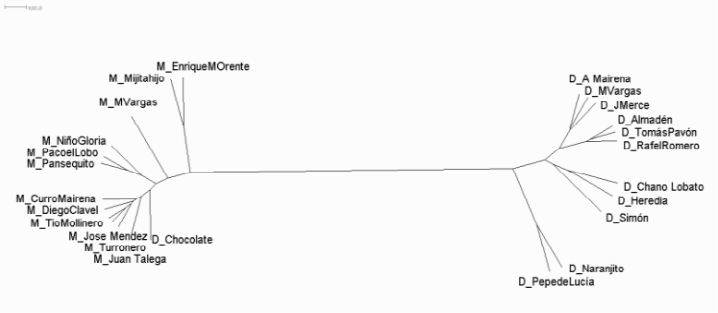


Figure 8: The SplitTree representation with the sub-styles deblas and martinetes.

3.0.2 Evalution methodology

For each of the system setups (S-I to S-VI) we generate the distance matrix and analyze the obtained clusters. In perfect classification scenario, two clusters would appear, corresponding to the two sub-styles. We evaluate the quality of the cluster separation by employing a kNN classification: An instance is correctly classified if the majority of the ground truth class of the k instances with the lowest calculated distance corresponds to its ground truth class. Figure 8 displays the phylogenetic graph for the fitting segmentation. Labels D (resp. M) stand for debla (resp. martinete).

We also analyze *precision* (Pr), *recall* (Re) and *f-measure* (FM) of the *martinete* examples being correctly retrieved. Furthermore, we assess the duration of the DTW calculations t_{DTW} as well as the total runtime t_T, both in seconds.

3.0.3 Results

The results shown in Table 1 indicate that for the given task an un-quantised note representation obtained the best overall classification results. Furthermore, the time needed for the alignment was significantly reduced when compared to the raw pitch contour. The curve fitting algorithm reduces the computation time, but leads to inferior classification

System setup	t_{DTW}	t_T	Pr	Re	FM
S-I	2290	2290	0.94	0.67	0.78
S-II	2290	2317	0.94	0.67	0.78
S-III	171	198	0.93	0.54	0.68
S-IV	244	245	0.94	0.63	0.75
S-V	5.48	8.48	0.96	0.96	0.96
S-VI	5.05	9.88	1.00	0.96	0.79

Table 1: Computational time and accuracy results.

results. We also observe that the vibrato removal does not lead to any improvement regarding both aspects discussed. The note quantisation does not improve the results. We suspect that this might be due to the unusable tuning during the performances.

4. CONCLUSIONS AND FUTURE WORK

Characterizing melodic similarity among pairs of strongly ornamented pitch contours is a complex task. Among other non-Western music traditions, we encounter such contours in a cappella flamenco singing. Nevertheless, efficient and accurate algorithms can provide the basis for a large-scale musicological analysis to gain deeper insight into intra-style categorization and evolution. We investigated a dynamic time warping approach to characterize melodic similarity and compared different pre-processing stages regarding the resulting accuracy and computational cost. We show that a musically meaningful transformation of the contour into piece-wise constant segments significantly improves both aspects. We are currently aim to improve the alignment method by incorporating intervals among adjacent notes instead of absolute pitches, since the latter leads to alignment errors when the pitch ranges of both sequences strongly differ. We furthermore target the development of an improved system for note-level transcriptions as well as to apply our approach with other flamenco styles.

Acknowledgments. This research was partly funded by the Junta de Andalucía, project COFLA: Computational Analysis of Flamenco Music, FEDER-P12-TIC-1362, the PhD fellowship of the Departament de Tecnologies de la Informaci i les Comunicacions (DTIC), Universitat Pompeu Fabra and the Spanish Ministry of Economy and Competitiveness (SIGMUS, Subprograma de Proyectos de Investigacin Fundamental no Orientada, TIN2012-36650).

5. REFERENCES

Cabrera, J. J., Díaz-Báñez, J. M., Escobar-Borrego, F. J., Gómez, E., Gomez Martin, F., & Mora, J. (2008). Comparative melodic analysis of a cappella flamenco cantes.

Díaz-Báñez, J. M. & Mesa, J. A. (2001). Fitting rectilinear polygonal curves to a set of points in the plane. *European Journal of Operational Research*, *130*(1), 214–222.

Díaz-Báñez, J. & Rizo, J. (2014). An efficient dtw-based approach for melodic similarity in flamenco singing. In *Similarity Search and Applications* (pp. 289–300). Springer.

Dressler, K. & Strech, S. (2007). Tuning frequency estimation using circular statistics. In *Proceedings of the 8th International Conference on Music Information Retrieval*, (pp. 357–360).

Goldstone, R. L. & Son, J. Y. (2005). *Similarity.* Cambridge University Press.

Gómez, E. & Bonada, J. (2013). Towards computer-assisted flamenco transcription: An experimental comparison of automatic transcription algorithms as applied to a cappella singing. *Computer Music Journal*, *37*(2), 73–90.

Herrera, P. & Bonada, J. (1998). Vibrato extraction and parameterization in the spectral modeling synthesis framework. In *Proceedings of the Digital Audio Effects Workshop (DAFX).*

Kroher, N., Gómez, E., Guastavino, C., Gómez-Martín, F., & Bonada, J. (2014). Computational models for perceived melodic similarity in a cappella flamenco cantes. In *15th International Society for Music Information Retrieval Conference.*

Molina, E., Barbancho, I., Gómez, E., Barbancho, A. M., & Tardón, L. J. (2013). Fundamental frequency alignment vs. note-based melodic similarity for singing voice assessment. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, (pp. 744–748). IEEE.

Mongeau, M. & Sankoff, D. (1990). Comparison of musical sequences. *Computers and the Humanities*, *24*(3), 161–175.

Mora, J., Gomez Martin, F., Gómez, E., Escobar-Borrego, F. J., & Díaz-Báñez, J. M. (2010). Characterization and melodic similarity of a cappella flamenco cantes. International Society for Music Information Retrieval Conference, ISMIR.

Myers, C., Rabiner, L., & Rosenberg, A. E. (1980). Performance tradeoffs in dynamic time warping algorithms for isolated word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *28*(6), 623–635.

Pikrakis, A., Theodoridis, S., & Kamarotos, D. (2003). Recognition of isolated musical patterns using context dependent dynamic time warping. *Speech and Audio Processing, IEEE Transactions on*, *11*(3), 175–183.

Sakoe, H. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Audio, Speech and Language Processing*, *26*, 43–49.

Salamon, J. & Gómez, E. (2011). Melody extraction from polyphonic music signals using pitch contour characteristics,". *IEEE Transactions on Audio, Speech and Language Processing*, *20*(6), 1759–1770.

USING PITCH FEATURES FOR THE CHARACTERIZATION OF INTERMEDIATE VOCAL PRODUCTIONS

Lionel Feugère, Boris Doval
LAM-Institut Jean Le Rond d'Alembert
CNRS UMR 7190, Sorbonne Universités
UPMC Univ Paris 06, F-75005, Paris, France
lionel.feugere,boris.doval@upmc.fr

Marie-France Mifune
UMR 7206 Eco-anthropologie et Ethnobiologie
CNRS-MNHN-Université Paris Diderot-
Sorbonne Universités, Paris, France
mifune@mnhn.fr

ABSTRACT

This paper presents some pitch features for the characterization of intermediate vocal productions from the CNRS - Musée de l'Homme sound archives, in the context of the DIADEMS interdisciplinary project gathering researchers from ethnomusicology and speech signal processing. Different categories – chanting, singing, recitation, storytelling, talking, lament – have been identified and characterized by ethnomusicologists and are confronted by acoustic analysis. A database totalizing 79 utterances from 25 countries spread around the world is used. Among the tested features, the note duration distribution has proved to be a relevant measure. Categories are mostly characterized by the proportion of 100-ms notes and the duration of the longest note. An evaluation of these features has been realized through a supervised classification using the different vocal categories. Classification results show that these two features allow a good discrimination between "speech", "chanting" and "singing", but are not suited for discriminating between the "speech" subcategories "recitation", "storytelling" and "talking".

1. INTRODUCTION

The context of this study [1] is the DIADEMS project [2], which gather together ethnomusicologists, linguists, acousticians, archivists and specialists in speech processing and music information retrieval, around a sound archive web platform, Telemeta (Fillon et al., 2014). This software platform allows the user to browse, listen, watch and annotate multimedia files.

The aim of the DIADEMS project is to develop computational tools to automatically index the audio content of the CNRS - Musée de l'Homme sound archives (49,300 audio items from 5,800 collections including 28,000 items already uploaded). This ethnomusicological archive includes published and unpublished recordings of music and oral traditions from around the world, spanning a wide variety of cultural contexts worldwide starting in the 1900s until today as well as a wide variety of contents (musical practice, speech, dance, ritual, interview and so on) and various settings (inside, outside, rarely in studio settings). Many sound archives in the collection include very little contextual information. The use of automatic indexation tools will help archivists to index such sound items and add new content information. It will also facilitate searches, analyses and comparison of large corpus by ethnomusicologists.

The core of the project deals with automatic indexing of a large variety of sound productions directly from audio signals including musical instruments, environmental sounds and vocal productions. In the context of the DIADEMS project, this study aims at helping ethnomusicologist to characterize speech, song and intermediate vocal productions in terms of acoustical parameters.

2. ETHNOMUSICOLOGICAL APPROACH

Several studies in ethnomusicology have characterized speech, song and intermediate forms through musicological and acoustic parameters in relation to the social and cultural context (Seeger, 1986; Amy de la Bretèque, 2010; Rappoport, 2005).

While the classic ethnomusicological approach considers that it is not possible to define vocal categories independently of the cultural context (Picard, 2008), some ethnomusicologists proposed attempts of classification of vocal productions among the different cultures.

The ethnomusicologist List (1963) proposed a method of classification by which distinctions and relations can be made between speech, song and intermediate forms. This classification system is based on two divergent modifications of speech intonation: 1) the negation or the leveling out of speech intonation into monotone, 2) the amplification or exaggeration of speech intonation (such as Sprechstimme). The next step to song is either 1) the stability of pitch or 2) the expansion of scalar structure. In his classification model, List didn't directly take into account the pitch duration. According to him, the comparative length or shortness of sustained pitch would be a useful criterion to incorporate in this system of classification but the assumption that song exhibits pitches of greater duration than speech would not be valid in all culture.

For particular productions like lament, Urban (1988) have found that some common features could be brought out among different cultures.

It is extremely challenging for ethnomusicologists to define an efficient categorization of vocal productions based only on acoustic criteria and equally efficient in all cultural practices worldwide for two main reasons: 1) procedures and techniques of vocal productions worldwide are insuf-

[1] This work is partly supported by a grant from the French National Research Agency (ANR) with reference ANR-12-CORD-0022.

[2] http://www.irit.fr/recherches/SAMOVA/DIADEMS/fr/welcome/

ficiently described and the inventory is incomplete; 2) ethnomusicological method and acoustical terminology often lack consensus and are somewhat approximate.

The first attempt of classification of vocal productions worldwide was a typology with audio examples based on several techniques (Zemp et al., 1996): calls, cries, clamours, voice and breath, spoken, declaimed, sung, compass and register, colours and timbres, disguised voices, ornamentation, voices and musical instruments, employ of harmonics. The main issue is that these several categories are not systematically based on evident and explicit acoustic criteria. Then, based on this classification, Léothaud (2007) put forward a universal typology of vocal techniques based on acoustic criteria such as timbre, register, tessitura and intensity.

As Giannattasio (2007) suggested, realizing a typology implies an analysis based on common acoustical parameters (tempo, intonation, timbre, etc.). By exploring the continuum between speech and song, we must define the different modalities of expression based on the several parameters without defining a predetermined order among them.

This study intends both to describe acoustical features that characterize the vocal categories and that would apply to different voice production excerpts from all over the world, and to draw up definitions of each vocal category from an ethnomusicological point of view.

We chose to classify vocal productions in two general categories: speech and song for two main reasons: 1) every culture distinguishes between talking and singing among all vocal productions; 2) we consider multiple vocal categories with acoustic characteristics ranging from speech to song.

Then, according to the database, we subsequently identified and characterized subcategories such as talking, storytelling, recitation, chanting, singing and lament, not based on style or genre, but on acoustic features only. We defined these subcategories without establishing a predetermined order among them. We define *speech* as a vocal production with a significant proportion of unvoiced sounds. Alternatively, *song* is defined as a vocal production with a significant proportion of lengthened syllables and voiced sounds. We consider *talking*, *recitation* and *storytelling* close to speech, since they are also characterized with a significant proportion of unvoiced sounds. We distinguish *talking* from *storytelling* based only on the mode of realization: *talking* is characterized by dialogue and *storytelling* by monologue with or without back-channel signal (i.e., an expression or word used by a listener to indicate that he or she is paying attention to the speaker). *Recitation* is characterized by more regular breath rate and rhythmic flow than *talking*, and a monotonous statement with low frequency range variations. We consider *singing* and *chanting* close to song, since they are characterized with a significant proportion of lengthened syllables and voiced sounds. *Singing* is characterized by ordered pitches and relative stability of fundamental frequencies while *chanting* is characterized by a very limited vocal range, close to recto-tono. We define *lament* by the presence of several of the four common icons of crying (the cry break, the voice inhalation, the creaky voice and the falsetto vowel) proposed by Urban (1988).

These definitions are a first attempt based on the ethnomusicological archives with which researchers in the DIADEMS project are most familiarized. Ultimately, the goal is to refine these categorical definitions while expanding the corpus of recordings considered for the automatic indexation. In particular, we are aware that some of the terminology used here can be inappropriate for some specific practices or can be confusing for some ethnomusicologists in the community. The aim of the definition and characterization of the ethnomusicological categories is not to replace the endogenous categories but to provide scientific tools to better analyze vocal productions. This work focuses more on the characterizations of descriptors and acoustic parameters rather than in the definition of the category themselves.

In the information retrieval community, few have addressed the issue of intermediate vocal categories, focusing more on cultural style (Liu et al., 2009), singing voice timbre (Fujihara & Goto, 2007), speech style (Goldman et al., 2009), or singing versus speech classification on homogeneous and/or good quality recordings (Gärtner, 2010).

Section 3 presents the corpus and the features that have been studied, section 4 presents the results in term of characterization and classification rates, and section 5 presents a discussion and some perspectives.

3. METHODS

3.1 Corpus

Ethnomusicologists from the Diadems project manually annotated the audio contents of a representative sample of sound items from the CNRS - Musée de l'Homme sound archives, in order to give to acousticians a data set of each subcategory mentioned above.

This resulted in 79 items from different contexts (rituals, enquiries, tales, etc.) and various cultures as shown in table 1. They were selected for their non-ambiguous category during 10 sec minimum. In each category, most of the utterances are from a different field. Except if indicated, the utterances from a same country and a same category are from a different speaker (but also often from a different context).

3.2 Previous study on intermediate vocal categories from the Telemeta corpus

In the context of the DIADEMS project, Sotiropoulos (2014) proposed a decision tree to classify utterances to four categories (chanting, singing, recitation and speaking/storytelling) from a small subset of the Telemeta database, composed of 6 utterances by category (including sometimes the same speaker). Mean voicing duration discriminates song and speech categories, while in the secondary nodes of the decision tree, mean duration of non-voice units divides talking+storytelling and recitation, and pitch range allows to

Class	Number	Origin (number)
Chanting	22	Cambodia (2), South India, Indonesia (7), Iran, Ladakh, Mexico (4), Nepal (3 including 2 same speakers), Tibet, Vietnam (2)
Singing	19	Albania, Armenia (2), Azerbadjan, Egypt, Gabon (2), Indonesia (3), Macedonia, Madagascar, Morocco, Nepal, the Philippines, Yemen (3), Turkey
Storytelling	8	Central African Republic, Gabon (5 including 2 same speakers), Mali, Paraguay
Recitation	8	Madagascar, Mali (3 including 2 same speakers), Mexico, Paraguay, Tibet, Yemen
Talking	12	France (2), Gabon (8), Mexico, Madagascar
Lament	10	Albania, Armenia, Azerbaijan, Ethiopia, Gabon, Paraguay (2 same speak.), Turkey(2), Vietnam

Table 1: Number and origin of the sound utterances used in this work.

separate singing and chanting. These features will be gathered with the present article ones in the future final system.

3.3 Features

3.3.1 Chroma spectrum

The audio recordings are from all over the world and many cultures, so no pitch reference is there and we are not necessarily in an equal tempered 12-semitones scale. Then instead of using chroma vector (i.e. 12 discrete notes) as it is mostly done (Harte & Sandler, 2005; Lartillot et al., 2007), we use chroma spectrum (Dannenberg & Goto, 2008).

The frequency axis of a spectrum is interpolated to get semitones units and the octaves are summed in order to have all the frequency axis information inside one octave. In the resulting chroma spectrum, the number of peaks (with an amplitude above a given threshold) and their width are computed.

3.3.2 Note distribution

Besides the audio recordings are not studio recorded, they are very often noisy, polyphonic and accompanied with other instruments. So we decided not to use standard F0 detection, but rather to design a quite simple but robust algorithm adapted for ethnomusicological voice recordings. We chose to detect all the partials (named note) with a sufficient energy in the spectrogram. For each note, its duration is determined and the resulting note duration distribution is computed for each 10-sec audio utterance.

First a modified spectrogram is computed as follows:

- If the audio is stereo, a mono signal is computed
- The spectrogram is built by computing the magnitude spectra every 50ms on 100 ms windows, over 10 sec of the audio file, and on a 5 octaves scale from 110 Hz to 3520 Hz.
- An interpolation is done in order to transform the frequency axis in log2 scale with 110 Hz as reference.
- Two thresholds are applied: a dynamic energy threshold equal to 1.5 times the mean energy computed on a local 500-ms window, and an absolute energy threshold computed from the noise level.

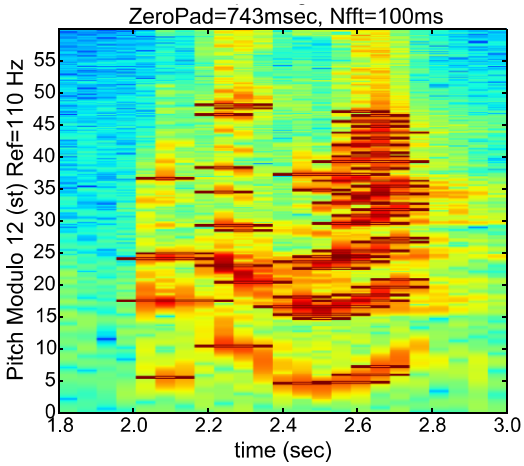


Figure 1: Spectrogram (frequency axis in semitone) annotated with the detected notes (in brown).

Then notes are detected, defined as energy on a constant bin frequency ± 0.175 semitones. The interval of ± 0.175 semitones is related to the tremor frequency range (Stables et al., 2012). For each frequency bin ± 0.175 semitones, energy greater than 0 is searched for from the initial bin window. A note is considered as finished if no energy is found for one time sample (50 ms) inside the frequency interval around the bin. Then, the note duration is computed. As an analysis step of 50 ms is used, the note duration can take values from 50 ms to 5 sec by steps of 50 ms. Finally, adjacent notes along the frequency axis are grouped together in order to avoid multiple note detection for a single energy peak. Figure 1 gives the detection result on an audio signal of 1.2 sec. The notes are surrounded by brown thin rectangles.

A distribution of the note durations is computed, normalized by the total duration of the detected notes. This normalization allows the distribution to be independent from similar repetitions of any vocal production.

Then each note duration proportion is multiplied by its note length, so that the resulted distribution can be thought of as a proportion in duration rather than a proportion in number of notes.

Figure 2 gives an example of two distributions: an audio file labelled as talking and another one labelled as chanting.

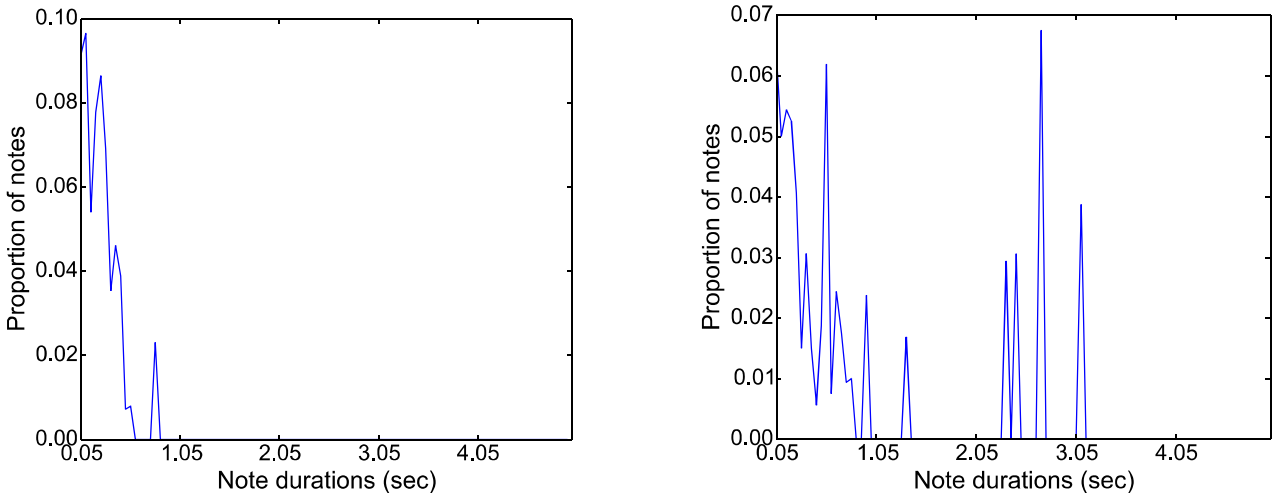


Figure 2: Note duration distributions for two excerpts. Left: talking. Right: Chanting.

Notice that no other vocal feature is taken into account to detect notes, like vowel or consonant articulation. As our note detection system focuses on pitch and energy only, the consecutive syllables on a same pitch are grouped into one single note, which occurs quite often in chanting (and spread in several utterances along the frequency axis). In this case, the number of long notes, as we defined it, is increased.

From this distribution, we chose to compute the note duration range (NDR), which is the longest note duration. On figure 2, the left distribution has a NDR of 0.9 sec while the right distribution has a NDR of 3.05 sec.

Other parameters are computed from the note detection. The normalized total duration of detected notes (named TotDurNote) is the total duration of detected note divided by the audio file duration. The mean instantaneous note number (named InstNoteNum) is computed by dividing the total duration of the detected notes by the segment duration where notes have been detected. The voicing proportion (named VoicProp) is the segment duration where notes have been detected divided by the audio file duration.

4. RESULTS

4.1 Characterization from the note duration distribution

Statistics on note duration distributions are displayed figure 3. Each figure is related to one category and displays the note duration distribution for all sound utterances of this category. For each note duration proportion, statistics over the utterances from a same category are represented by the help of the following tools: a vertical box for which borders correspond to the first and third quartile (i.e. quarter of the utterances are above the box while quarter of the utterances are below the box, so half of the data is included in this box); an horizontal red line showing the median (i.e. half of the data are spread above the line, the other half below the line); vertical whiskers representing 50% of the data plus 3 times the interquartile range (third quartile minus first quartile); fliers representing data that extend beyond the whiskers (outliers).

The main difference in the note duration distribution between categories is the proportion of very short and long notes.

For each category, table 2 gives the note duration range NDR (calculated either from the median or from the outliers) and the coordinate of the median maximum of the note duration distribution. Notice that the different dispersions between the categories are related to their number of utterances, which are quite different in our database.

4.2 Characterization from 100-ms note proportion and note duration range

Several features computed from the note detection are discriminant but the result is different from one category to another and it depends on the number of considered categories. One of the most discriminant couple of features for the whole categories are the 100-ms note proportion (second value of the note duration distribution) and the note duration range (NDR), using cross validation protocol and k-Nearest Neighbour classifier, so they were chosen to display the utterances.

Displayed in the 2D space formed by the values of these two parameters, the utterances are expected to group into characteristic areas related to their category.

As shown on figure 4:

- singing category is characterized by a small value of note duration range and proportion of 100-ms notes;
- chanting category is rather characterized by a small value of proportion of 100-ms notes and a large value of note duration range;
- speech categories (talking, recitation and storytelling) are characterized by a small value of note duration range and a large proportion of 100-ms notes;
- lament category is characterized by a small proportion of 100-ms notes.

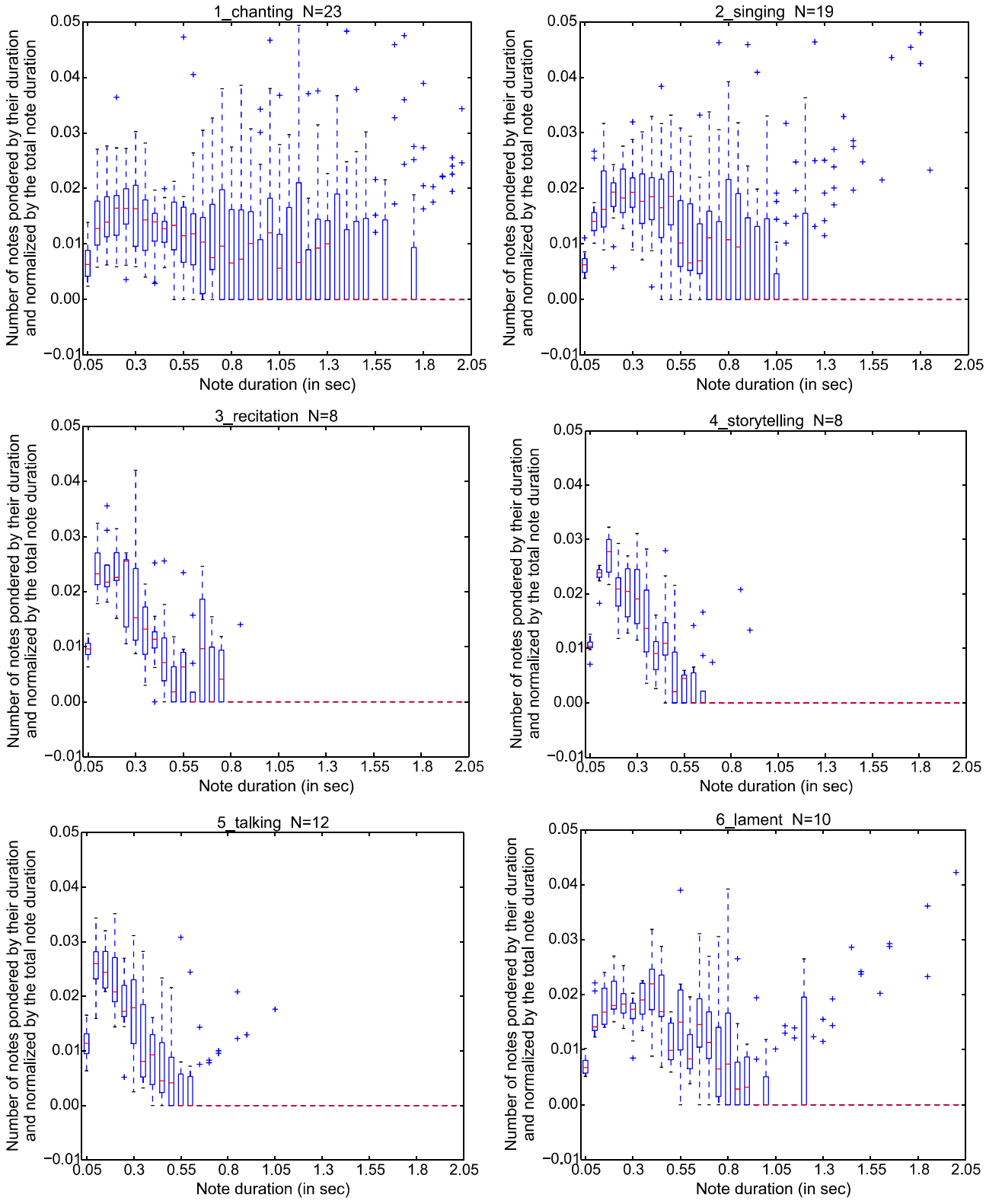


Figure 3: Median values and dispersion of the note duration distribution for the 6 categories chanting, singing, recitation, storytelling, talking and lament (see the text for more information).

Parameter	Chanting	Singing	Recitation	Storytelling	Talking	Lament
NDR (median)	1.30 sec	0.85 sec	0.75 sec	0.55 sec	0.50 sec	0.90 sec
NDR (all points except outliers)	1.75 sec	1.20 sec	0.75 sec	0.65 sec	0.60 sec	1.20 sec
Distrib. maximum (median)	0.0164	1.93%	2.55%	2.78%	2.60%	2.20%
and duration	0.20 sec	0.20 sec	0.25 sec	0.15 sec	0.10 sec	0.40 sec

Table 2: First line: note duration beyond which the median is 0. Second line: note duration beyond which non-0-values come from outliers only. Third line: coordinate of the median maximum of the note duration distribution.

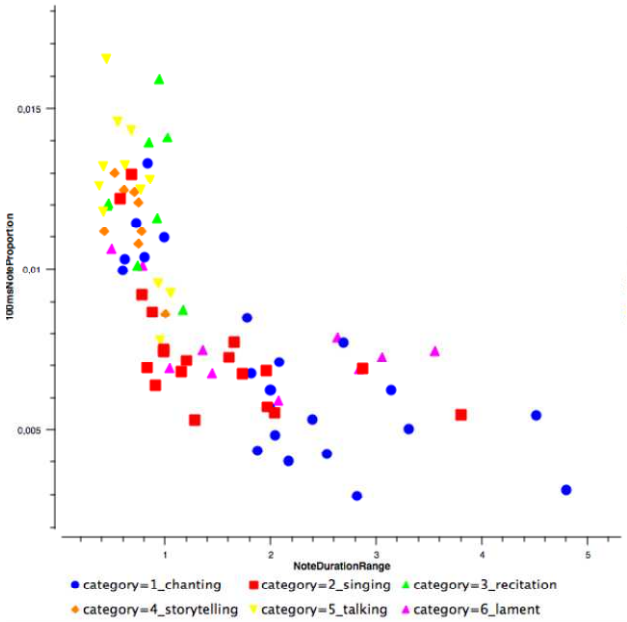


Figure 4: Utterance position in a 2D space according to their values of NDR and to 100-ms note proportion.

As can be seen on figure 4, the items from talking, storytelling and recitation are mostly overlapped, the items from chanting and singing are partially overlapped and lament items are mostly overlapped with singing or chanting ones. Then it seemed interesting to study the results when grouping speech categories from one side (talking, storytelling and recitation), and song categories from the other side (singing and chanting), as displayed in figure 5. Notice that lament was not considered as a distinct category by ethnomusicologists.

These overlappings are quite easily explained. First, on the 3-classes figure, the two singing utterances [3] which have a large value of NDR have long notes like in chanting, while the three singing utterances [4] with high value of proportion of 100-ms notes are very articulated like in speech. Second, on the same figure, the 6 chanting utterances [5] with a low value of NDR are the most articulated among the chanting instances. Third, the 4 chanting utterances [6] with the smallest values of NDR and 100-ms note proportion are somehow quite close from their neighbour utterances in term of note durations. Lastly, a long syllable is found in the bottom-right speech utterance [7] with the greatest note duration range.

4.3 Evaluation by supervised classification

In order to evaluate quantitatively our features, supervised classification were performed with the database and the best classifier was selected [8]. Results are reported below in the form of confusion matrix with different groupings among the 6 vocal categories. The rows correspond to true categories while the columns are the predicted categories. For instance, in table 3, the utterances labelled as chanting by the ethnomusicologists were correctly classified by the algorithm as 57%, while 22% were classified in the singing category and none was classified in the lament (0%).

Classification was done using the following features: note duration distribution, note duration range, normalized total duration of detected notes (TotDurNote), mean instantaneous note number (InstNoteNum), voicing proportion (VoicProp), peak number and peak width of the chroma spectrum.

First, classification with all the 6 categories gives some quite bad results (see table 3), including the storytelling class with a true prediction score less than with a random classification (for 6 classes: 17%) and the others between 25% and 57%.

As the speech classes are not well discriminated between each other, we grouped the talking, storytelling and recitation together to form the speech class, which is detected with a 82% rate (for a total of 28 utterances) against the chanting and singing classes (see top table 4). The chanting class gets a 65% score of true prediction while singing class performs 74% of true detection.

If grouping singing and chanting categories as one, named song, then about 9/10 of the sound examples are well classified (bottom table 4).

If considering song, speech and lament as three categories to be classified, lament are classified as song for 90% of their utterances, 10% as speech, and 0% as lament

[3] 10 first seconds of CNRSMH_I_2012_004_001_002 and CNRSMH_E_1992_007_002_001_11

[4] 10 first seconds of CNRSMH_I_1971_025_003_020_12s-39s, CNRSMH_I_2003_010_001_04, CNRSMH_E_1959_002_004_003_02_1mn14-end

[5] 10 first seconds of CNRSMH_I_1970_068_016_01_30s-end, CNRSMH_I_1975_015_017_05, CNRSMH_I_1972_013_025_02_1s-end, CNRSMH_I_1975_015_017_02, CNRSMH_I_1972_012_015_05_6s-end, and CNRSMH_I_1972_012_015_03

[6] 10 first seconds of CNRSMH_I_1970_068_010_02, CNRSMH_I_2011_016_008_04_37s-end, CNRSMH_I_1970_071_009_04_55s-end, CNRSMH_I_1959_006_001_01

[7] 10 first seconds of CNRSMH_I_2007_005_044_01_30s-end, CNRSMH_I_1974_003_006_01_2s-end

[8] Among Naive Bayes, Classification Tree, SVM, kNN, Neural Network, Random Forest, CN2 rules.

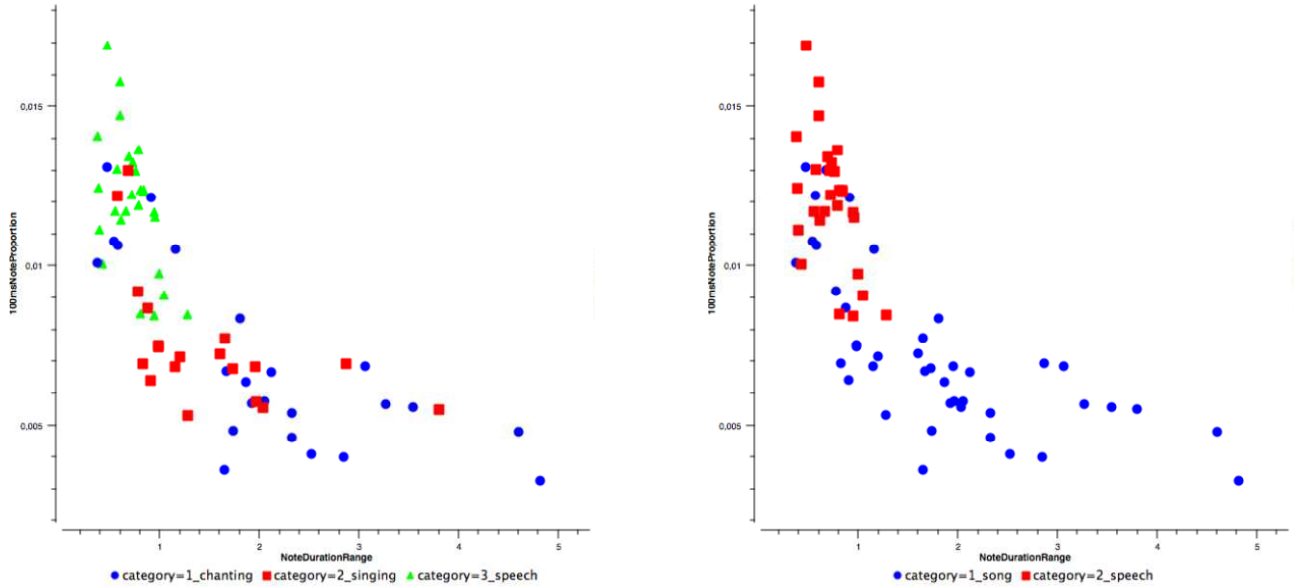


Figure 5: Utterance position in a 2D space according to their values of NDR and 100-ms note proportion. Left: Chanting VS Singing VS Speech (i.e. recitation+storytelling+talking). Right: Song (chanting+singing) VS Speech.

	Chanting	Singing	Recitation	Storytelling	Talking	Lament
Chanting	57%	22%	13%	4%	4%	0%
Singing	5%	53%	0%	0%	16%	26%
Recitation	50%	12%	25%	0%	13%	0%
Storytelling	12%	0%	25%	13%	50%	0%
Talking	25%	17%	0%	17%	33%	8%
Lament	20%	30%	0%	0%	10%	40%

Table 3: Confusion Matrix of the 6-classes classification (proportions of true) using a Classification Tree and a cross validation procedure. Columns represent predictions, rows represent true classes.

while song and speech respectively reach a score of 79% and 89%. On the whole, grouping song and lament categories improves the true prediction score: 90% for the category song+lament and 86% for speech. It means that our lament category is closer to song than speech.

	Chanting	Singing	Speech
Chanting	65%	18%	17%
Singing	16%	74%	10%
Speech	0%	18%	82%

	Song	Speech
Song	86%	14%
Speech	7%	93%

Table 4: Confusion Matrix (proportions of true) of the 3-classes (resp. 2-classes) classification using a Naive Bayes (resp. Random Forest) learner and cross validation. Columns represent predictions, rows represent true classes.

5. DISCUSSION AND PERSPECTIVES

We studied some features computed mostly from note duration distribution, and we discussed the characterization of categories by the proportion of 100-ms note and the note duration range. Nevertheless, the discrimination is reached with the help of the whole features, including all the note duration distribution and additional pitch features like voicing proportion, normalized total duration of detected notes, and mean instantaneous note number. Chroma features were found relatively non-discriminating.

The results show that these features are useful for classifying singing, chanting and speech but not for discriminating speech categories (storytelling, recitation, talking). The lament category, chosen for the presence of icons of crying, was found to be closer to song than to speech. Our study confirm the previous work by Sotiropoulos (2014) that it is possible to group vocal productions in acoustically coherent categories.

From an ethnomusicological point of view, these results bring new perspective on the definitions of vocal categories. Classic ethnomusicological approach focuses on endogenous categorizations of musical practices, thus specific to each culture and never solely based on acoustic criteria. The DIADEMS project ambitioned to build a transver-

sal characterization of vocal categories sampled from different cultures and based only on acoustic parameters.

The aim of this study is to test acoustic parameters on several scientific categories defined and characterized by ethnomusicologist. This work focuses more on the characterizations of descriptors and acoustic parameters, rather than on the definition of the categories themselves.

Results show that talking, storytelling and recitation categories can not be distinguished based only on the acoustic features here tested. Results support the definitions given by ethnomusicologists: talking and storytelling only differ by the mode of realization (monologue/dialogue), which is not embedded in pitch features. Furthermore, results corroborate the distinction made by ethnomusicologists between speech and song and their respective subcategories (talking, storytelling, recitation concerning speech and singing, chanting concerning song).

The results show that the lament category seems closer to song based on pitch features. This classification does not support the ethnomusicological definition proposed by Urban but brings new acoustic characterization. To test the lament category and the icons of crying proposed by Urban, other parameters than note duration distribution should be tested. Results may be also biased by our test dataset and must be further tested using other datasets.

Concerning implementation, this system will be improved and completed by other features from the literature (Sotiropoulos, 2014) and other type of pitch features, especially taking into account the time evolution of the detected notes. Our work is intended to be included in the timeside library [9], an open source plugin used in the Telemeta interface as a graphical help for ethnomusicologists. However, the final user interface in Telemeta should not give predicted category names only, but rather more subtle raw informations to help the ethnomusicologists with the data indexation.

6. REFERENCES

Amy de la Bretèque, E. (2010). Des affects entre guillemets. mélodisation de la parole chez les Yézidis d'Arménie. *Cahiers d'ethnomusicologie*, *23*, 131–145.

Dannenberg, R. B. & Goto, M. (2008). *Handbook of Signal Processing in Acoustics*, chapter Music Structure Analysis from Acoustic Signals, (pp. 305–331). Springer New York.

Fillon, T., Simonnot, J., Mifune, M.-F., Khoury, S., Pellerin, G., Le Coz, M., Amy de la Bretèque, E., Doukhan, D., & Fourer, D. (2014). Telemata: An open-source web framework for ethnomusicological audio archives management and automatic analysis. In *Proc. 1st Digital Libraries for Musicology workshop (DLfM 2014)*, London, UK.

Fujihara, H. & Goto, M. (2007). A music information retrieval system based on singing voice timbre. In *8th International Society for Music Information Retrieval Conference*.

Gärtner, D. (2010). Singing / rap classification of isolated vocal tracks. In *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, (pp. 519–524).

Giannattasio, F. (2007). *Musiques. Une encyclopédie pour le XXIe siècle. L'unité de la musique* (Actes Sud/ Cité de la musique ed.)., volume 5, chapter Du parlé au chanté: typologie des relations entre musique et texte, (pp. 1050–1087). Arles, Paris.

Goldman, J.-P., Auchlin, A., & Simon, A. C. (2009). Discrimination de styles de parole par analyse prosodique semi-automatique. In *Actes d'IDP*, (pp. 2114–7612)., Paris.

Harte, C. & Sandler, M. (2005). Automatic chord identification using a quantised chromagram. In *Audio Engineering Society Convention*, volume 118.

Lartillot, O., Toiviainen, P., & Eerola, T. (2007). A matlab toolbox for music information retrieval. In *Proc. of International Conference on Digital Audio Effects*, Bordeaux.

Léothaud, G. (2007). *Musiques. Une encyclopédie pour le XXIe siècle. L'unité de la musique* (Actes Sud/ Cité de la musique ed.)., volume 5, chapter Classification universelle des types de techniques vocales, (pp. 803–832). Arles, Paris.

List, G. (1963). The boundaries of speech and song. *Ethnomusicology*, *7*(1), 1–16.

Liu, Y., Xiang, Q., Wang, Y., & Cai, L. (2009). Cultural style based music classification of audio signals. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, (pp. 57–60).

Picard, F. (2008). Parole, déclamation, récitation, cantillation, psalmodie, chant. *Revue des Traditions Musicales des Mondes Arabe et Mediterraneen (RTMMAM)*, 8–24.

Rappoport, D. (2005). Les langues frétillantes. Modalités de profération de la parole rituelle chez les Toraja d'Indonésie. In *Second Congress of Asia Network*, Paris. EHESS.

Seeger, A. (1986). *Native South American Discourse* (Mouton de Gruyter ed.)., chapter Oratory is Spoken, Myth is Told, and Song is Sung: But They Are All Music to My Ears, (pp. 59–82). Berlin.

Sotiropoulos, T. (2014). Caractérisation des voix intermédiaires : de la taxinomie des voix dans un contexte ethnomusicologique. Master's thesis, Université Paul Sabatier, équipe SAMoVA.

Stables, R., Athwal, C., & Bullock, J. (2012). *Speech, Sound and Music Processing: Embracing Research in India*, volume 7172 of *Lecture Notes in Computer Science*, chapter Fundamental Frequency Modulation in Singing Voice Synthesis, (pp. 104–119). Springer Berlin Heidelberg.

Urban, G. (1988). Ritual Wailing in Amerindian Brazil. *American Anthropologist*, *90*(2), 385–400.

Zemp, H., Léothaud, G., Lortat-Jacob, B., Tran, Q. H., & Schwarz, J. (1996). *Voices of the World. An Anthology of Vocal Expression*. Collection CNRS-Musée de l'Homme / Le Chant du Monde CMX 374 1010.12.

[9] https://github.com/yomguy/TimeSide/

Analysis of Brazilian popular music: the Pixinguinha arrangements for the orchestra's of the radio program *O Pessoal da Velha Guarda*

Ana Lúcia Fontenele
Universidade Federal do Acre
Universidade de São Paulo
alfontenele@usp.br

ABSTRACT

The purpose of this work is to describe the musical practices in the late 19th and early 20th centuries which have been restored in arrangements produced by *Pixinguinha* for the radio show program called *O Pessoal da Velha Guarda*, broadcasted by *Radio Tupi do Rio de Janeiro* between 1947 and 1952. In respect to the historical contextualization of Brazilian popular music it can also be observed the emergence of modern *samba* in Brazil as well as its influence on *choro* from the 1930s.Lastly, the main musicological features of the arrangements mentioned before were analysed through musical key points of Brazilian music suggested by Piedade (2013), especially in regard to the melodically and rhythmic aspects.

1. INTRODUCTION

The purpose of this study is to describe and analyse part of practises and styles of the music present in the city of Rio de Janeiro in the late 19th and early 20th century. Such musical practises began as a reinterpretation of some European ballroom dancing, especially polka and waltz, both at aristocratic halls and popular neighbourhoods in town. In the world of instrumental music these experiences promoted the appearance of *choro* and established the grounds of a great part of Brazil's urban popular musicality. Such musical styles were subsequently used by Pixinguinha for the orchestra of the radio program called "O Pessoal da Velha Guarda" which was broadcasted by Radio Tupi of Rio de Janeiro between 1947 and 1952.The purpose of the musical producer *Almirante*, and his fellow musicians, was to recover works performed both in noble halls and popular parties in the city of Rio de Janeiro.

On the other hand, we want to assert what was previously stated that "in spite of dialogues held with different musical styles such as *jazz*, *Pixinguinha* was faithful to the musicality practised in Brazil since late 19th century" (Fontenele, 2014).

This influence can be noticed in a more mature stage of *Pixinguinha's* career which started with the performance of arrangements for a Brazilian popular orchestra in the late 1940s.

Since 1980s some attempts to formulate analysis methods of popular music have established new perspectives based on analysis proposed by the New Musicology - it takes into account cultural and social aspects in which musical practises emerge from popular urban music.
In turn, in spite of being influenced by social and cultural aspects in their creative journey, composers and singers express themselves in a creative manner of their own.

In this sense, the social environment where these musical events happened as well as the names and relevant facts of some composers and singers of that time were minutely studied to provide a contextualized analysis of the orchestral arrangements previously mentioned.

1.1 The beginnings of urban popular music

In Brazil some musical styles such as *lundu, maxixe, choro* and *samba* emerged as a result of musical practises performed in social gatherings and parties of popular classes. Before they emerged, *batuques* from people of Africa and its descendants along with European dances such as waltz, polka, schottisches and quadrilles were always present at upper class balls and also adapted to popular parties.

In this process of cultural fusion, polka was the most influenced style by Brazilian musicality. Polka arrived in Brazil around 1844 and it is a ballroom dance originated in Bohemia with a strong duple rhythm. It was present in traditional balls including *Carnaval* and it was also present in some parties such as *Bairro Cidade Nova* in the centre of Rio de Janeiro. According to some writers such as *Tinhorão* (2010) polka became so Brazilian that originated *lundu, Brazilian tango, maxixe, choro* and *samba amaxixado.*

Distinguished people from the times of ballroom dancing were considered the pioneers of Brazilian instrumental popular music. Among them stands out the composer *Henrique Alves de Mesquita, Joaquim Callado,Viriato Figueira,* the Belgium songwriter *M.A. Reichert, Ernesto Nazareth and Chiquinha Gonzaga*, followed by *Irineu de Almeida* and *Mário Álvares da Conceição* (both teachers of Pixinguinha) and by *Anacleto de Medeiros and Luiz Souza.*

According to Severiano (2008) the Brazilian tango, previously mentioned as a hybrid style, results from the "mixture" between *habanera* and the Spanish tango along with some elements of polka and *lundu. Henrique Alves de Mesquita* created the Brazilian tango and Ernesto Nazareth was the one who best coordinated this style in a moment when syncopated genres were already acclaimed in Brazilian musical culture Machado (2007, p. 40).

Both *Nazareth* and *Viriato Figueira* (flautist) fellows of *Joaquim Callado* and *Patápio Silva* composed so refined polkas that challenged the possible frontier between popular and classical. Machado (2007) mentioned the dialogue *"Caiu não disse?"* in a Polk written by Viriato Figueira and Enesto Nazareth who answered with another Polk called *"Não caio n'outra!"* As per Tinhorão (1991,

p.15) *choro* singers had a virtuous concern about introducing unexpected modulations due to "bring down" the musicians.

The use of syncopation to interrupt the regular flow of rhythm[1] is present in subgenres of polka such as Brazilian tango and in the music of *Nazareth.* This use was highlighted both by *Mario de Andrade* and by the pianist *Arthur Rubinstein* who was delighted with *Nazareth*'s interpretations when the pianist stayed in Rio de Janeiro in 1918 (Machado, 2007).

At the time the French composer Darius Milhaud was in Brazil instrumental music was not only performing in fancy balls and popular neighbourhood's parties, but also in the withdrawing rooms at projection rooms. The composer created an orchestrated piano piece called *Saudades do Brasil* and the ballet *Le Boeuf sur Le toit* inspired by the music he listened in Brazil in the period between 1918 and 1919. According to *Corrêa do Lago* (2012) the ballet was composed in Paris in 1919 and includes more than a dozen pieces of Brazilian popular music whose lyrics he obtained from musical scores.

There was a kind of social arrangement regarding the people who often visited the ballrooms at the time of the second empire and the beginning of republic in Brazil (1922 to 1989) they were mid-class young women and men who played the piano and low class musicians who contributed to the formation of musical tradition. In the context of the parties of lower classes, instead of brass orchestras and pianos from elite balls, the "Brazilian style" European dances (waltz, polka, schottisches) and Brazilian styles such as Brazilian tango, *maxixe, partido alto and samba amaxixado* were performed by the groups of "wood and string", composed by flute, *cavaquinho* and acoustic guitar instead of metal orchestras or pianos.

Musicians, composers and interpreters identified these aspects and *Alexandre Gonçalves Pinto* mentioned them in the book: O *Choro: reminiscências dos chorões antigos* published in 1936 and republished in 1978 and 2009. *Alexandre* portrayed in his book called *O Animal* how much this musical practise was open to different interactions among people, listeners and *choro* practise singers:

> People look back to that period in Rio de Janeiro and feel in their souls the vibrations of the music from that time: The *chorões* at moonlight, the balls at family houses, and the ordinary parties where there was truthfulness, joy, hospitality and shared ideas (Pinto, 2009, p.10).

It is interesting to note that *Aragão* (2013) emphasizes about the meaning of "roda" related to the *choro* of that period (late 19th century and early 20th century): "the term *roda* was used as a synonym for the music society and *choro* lovers. This community according to the speech of the postman (*O Animal*) was connected by a sense of identity provided by determinant sound and social practises and also a past, a common tradition. In this sense there is in the text of *Alexandre Gonçalves Pinto* (2009) a great emphasis on musicians named as belonging to an "old guard".

1.2 African influence

The European ballroom dancing as polka, waltz, schottische and subsequently, *choro* and *samba* were influenced by elements of African music rhythmic present in music events of black people in Brazil such as *rodas de batuque* and others. As per Severiano (2008, p.69) *samba* would not exist if there were not different forms of folkloric *samba* and practise of *rodas de batuque* before.
Some factors contributed for a greater presence of black people in the capitals and at the end of the 19th century it happened mainly at the city of Rio de Janeiro, capital of Brazil at that time. According to Severiano (2008) the key factors were : the decline of coffee plantation, the end of the Paraguay War (1887), the end of the Canudos War (1897) and the North-eastern drought that occurred in 1877 until 1879 (Tinhorão, 1991, 2010 e Severiano, 2008).

In addition to styles as *modinha, lundu,* Brazilian tango and *maxixe*, the hybrid styles emerged in this process. It is worth highlighting the tango-maxixe, tango-lundu, and tango-batuque, among others. It is also important to note some references of theoreticians and scholars of urban cultures. *Garcia Canclini* recognizes aspects resulting from different cultural realities of a specific people and present in many countries of Latin America. This sense of cultural diversity was converted into an identity explicative model in a field where particular identities coexist (Garcia Canclini, 2006, p. 174 *apud* Carvalho, 2014).

1.3 Musical context in modern Brazil

In Brazil the industry of entertainment is established in an overwhelming manner because of the birth of the radio in 1922, the electric recording and the sound films in 1927. Such events changed the working dynamic of musicians and transformed the selection process of professionals more demanding than before. Consequently many composers from earlier periods were overshadowed and only some distinguished instrumentalist of high technical level remained active and performed on radio shows and albums of that period.

Around 1930 there was a sort of modernization of *samba* with the standardization of a kind of rhythmic accompaniment created by *sambistas* from *Estácio* neighbourhood in Rio de Janeiro[1], pointing according to Vianna (1955) for the samba homogenization idea around a single music aesthetics for the country (Figures 1 and 2). This phenomenon occurs when the samba becomes an eminently urban character of music, practised by blacks and whites in the process of becoming a synonym for national identity supported by the discourse of Brazilian intellectuals who put the samba as a positive reflection of Brazilian miscegenation. In this context according to Barata, "black culture could even form the identity of the

[1] Regarding the presence of the syncopa in European music, according to Machado (2007), it happened within the metric predictability of each measure , already in Brazilian music the syncopa happened , for example in samba, in a rhythmic series of two binary measures.

Brazilian nation but knowing its place, adapting and civilizing itself (Barata, 2012, p. 1796).

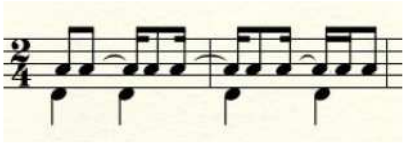


Figure 1. Rhythmic pattern of modern *samba* (Sandroni, 2012)

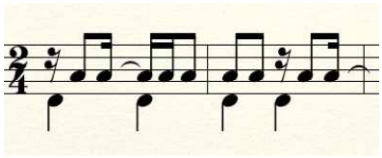


Figure 2. Rhythmic pattern variation of modern *samba* (Sandroni, 2012)

At this point regional groups that performed both lyric and instrumental songs emerged. *Choro* had a different shape especially in regard to the instruments for accompaniment. This new style is connected with the *samba* of *Estácio.* Among these regional groups the *Regional Benedito Lacerda* and the *Regional do Canhoto* stand out as they incorporated the characteristics of modern *samba* into *choro.* In addition, musicians from a new generation also introduced these new characteristics, among them some stand out, according to *Severiano* (2008): the composers *Ary Barroso, Noel Rosa, Braguinha, Ismael Silva*; the interpreters *Almirante* and *Mário Reis* and the instrumentalists as well as the arrangers *Radamés Gnattali, Benedito Lacerda, Carolina de Menezes, Dante Santoro, Luís Americano, Garoto, Rogério Guimarães, Meira, Luperce Miranda and Luciano Perrone* among others.

In an attempt to build a musical expression mass-focused the music arrangement is a key factor for the accomplishment of the project. Aragão (2001) states that the arrangement is an inherent process to any commercial popular music. The author asserts that:

> The arrangements used to fulfil the function of repackaging the music that came from lower classes into a noble fashion; however this "product" could not be presented in its raw shape as usual (Aragão, 2001, pg. 29).

As per *Trotta* (2009, pg. 2) the complexity o orchestral arrangements were meant to upgrade the practise of *samba* that is still waiting for social legitimating. In this context of local identity, it is transformed into a whirlpool of collages and hybridism and submitted to market demand (Martin-Barbero, 2005*, apud* Carvalho, 2014).

Apart from a few foreign arrangers performing in Brazil some popular Brazilian composers developed their art of creating arrangements; among them it is important to emphasize *Pixinguinha* and *Radamés Gnattali.* Since the 1930s until mid 1940s these two arrangers left their mark on the music business consumed in Brazil. It is worth stressing the great influence of North American music in the *big bands* established in Brazil from the 1940s. The *Tabajara* orchestra conducted by the musician, conductor and arranger *Severiano Araújo* stands out.

Despite those trends imposed by the record industry, *chorões* continued playing popular instrumental musical as *choro* with polka accompaniment which was considered by Aragão (2013, p. 119) the ultimate expression of what was known as "national music". According to Tinhorão (1997, p. 50), *samba* originated at *Estácio* in the early 1930s is up to today worshiped by *samba* interpreters from schools of *samba.*

On the whole, *carnaval samba*, *chorinho* and *marcha* continued developing taking into account these popular features originated in Rio de Janeiro. If in one hand the required high level of musicians has overshadowed the practising of a more free style of music, on the other hand this same high technical requirement lead to the appearance of virtuosos as *Jacob do Bandolim* and *Altamiro Carrilho,* among others. Accordingly *choro* has become well known and practised, not only in Brazil, but mostly in France, in the United States and Japan.

1.4 *The radio program: "O Pessoal da Velha Guarda."*

Since the 1940s, it is important to note some efforts to bring back the musicality that used to be one of the pillars of instrumental popular music in this respect the work of researchers as *Luiz Heitor Correa de Azevedo*[2]*, Brasílio Itiberê*[3]*, Mello de Moraes*[4]*, Almirante* and the bandolin player and composer *Jacob do Bandolim* among others. From 1955 to 1959 *Jacob do Bandolin* put together a show on the *Radio Nacional do Rio de Janeiro and* two television shows at *TV Record* in São Paulo gathering together a great number of regional[5] *choro* instrumentalists in the years of 1955 and 1956.

The composer and arranger *Pixinguinha* performed with Almirante as music director for the radio program *O Pessoal da Velha Guarda* (Paes, 2012). *Pixinguinha* is considered as the musical heir of musical practises connected with *choro* since the beginning of this music style because he used to work with composers and musicians from a previous generation such as, his teachers, *Irineu de Almeida* and *Mário Álvares, Pixinguinha* is considered as the musical heir of musical practises connected to *choro* since the beginnings of this music style especially characterized as the "Brazilian" fashion of playing dances originated in Europe. The composer wanted to recover the sounds of that musical experiences through the use of arrangements especially elaborated for the orchestra's radio program called *O Pessoal da Velha Guarda.* The arrangements created by him for the orchestra of the program were published in scores (Paes Leme, 2010 and 2014).

As stated before (Fontenele, 2014), *Pixinguinha* was considered as the great pioneer of orchestration for Brazilian popular music (Cabral, 2007), especially because he would adapt *choro* and samba instruments to the harmonics and rhythmic of this orchestra. He introduced the acoustic guitar and *cavaquinho* to play along with piano and acoustic bass and introduced some new instruments to percussion such as *cuica,* tambourine and *pandeiro* to be played along with the drum *jazz* band. The "Brazilian" popular orchestra had brass instruments (trumpets, saxophone and trombones), woodwind instruments (flute, pic-

colo, and clarinet) and stringed instruments composed mainly by violins (up to 4 *devisis*) in addition to the instruments of harmonic and rhythmic bases.

2. MUSICAL ANALYSIS

The "topics" theory adapted to the Brazilian popular music context was used by suggestion of *Acácio Piedade* (2013), to analyse the series of arrangements produced by Pixinguinha for the O *Pessoal da Velha Guarda* orchestra. The topic "brejeiro" is related to the melodic flourishes (portamento and appoggiatura) improvised mainly by soloists that were present in the arrangements of the main melody (Figure 3) and yet to anticipations of downbeat leading to irregular breaking and displacement of the rhythm (Figure 4). According to *Piedade* (2007) this kind of musical behaviour refers to the *malandro* figure, "smart and competent", with a swinging rhythm and a challenging and malicious spirit of "brejeira" topic.

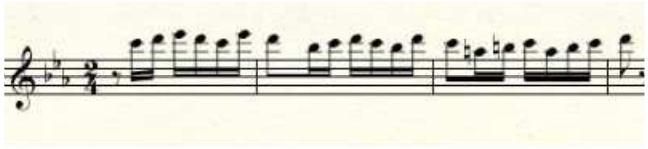


Figure 3. Melodic passage (1) from polka *"Assim é que é"* by Pixinguinha. (Paes Leme, 2010).

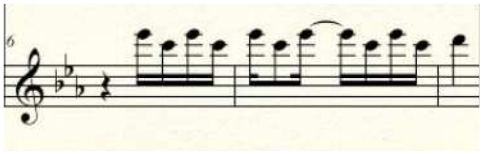


Figure 4. Melodic passage (2) from polka *"Assim é que é"* by Pixinguinha. (Paes Leme, 2010).

The topic "golden age" refers not only to the mannerisms in the melodic flourishes present in old Brazilian *serestas* waltz, but also in the different "*levadas*" of each style as a consequence of European dances that became more "Brazilian" and also in Brazilian music styles as *maxixe*, Brazilian tango and *samba*. (Figures 5 and 6).

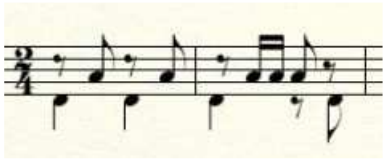


Figure 5. Rhythm pattern (1) from polka *"Assim é que é"* by Pixinguinha. (Paes Leme, 2010).

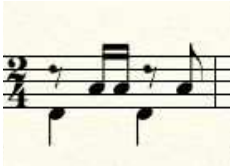


Figure 6. Rhythm pattern (2) from polka *"Assim é que é"* by Pixinguinha. (Paes Leme, 2010).

This perspective of analysis comes to an agreement with another proposition suggested by *Piedade* (2013) concerning an interaction of topics in a predictable field supported by the predictably that such musical elements bring to a composition or a musical arrangement. The argumentation on "commonplace" turns the musical flow into an organic experience. When there is a figure that stands out in this isotropic field of musical expression, a topic emerges.

In regard to *choro* the harmonic and formal aspects are related to what *Piedade* describes as "commonplace." Harmonic treatment inherited from tonal classical music can be analysed from the perspective of functional harmonic. Therefore, the analysis may continue from the melodic profile in major and minor tonalities and harmonic progressions extended to the field of "chromatic achievement" Webern (1984) because of the use of a dominant over each scale, the "secondary dominants", present in vocal music since the choirs of Bach and other kind of chords with auxiliary functions.

Also in the context of *choro,* the form, in general, results from minuets (ternary form) and most frequently, the rondo form (ABACA), where each musical part is in a different tonality. The constant return to part A may be considered as the highest level of the topic "commonplace". In the context of concert music, Webern (1984) considers modulation as the last stage of tonal music, especially those that occur in the form of *allegro* sonata.

3. DISCUSSION

After years of experience in the Brazilian music market as arranger, *Pixinguinha,* in its mature phase of his career, back to the kind of musicality that formed him. In a certain extent, he never abandoned the music practiced at the time the emergence of *choro* genre.

Another type of musical aesthetics, which was influential the arrangements of Pixinguinha, was the military bands, whose greatest representative is the composer *Anacleto de Medeiros*.

In their arrangements of this time and in previous phases, where Pixinguinha acted as arranger in belonging to discs record groups, such influences were present in his creations.

Among them there is the recording of some of Pixinguinha own albums released in the sixties and seventies of last century, as the album *Pixinguinha e sua Gente* - 1972- *Em tempo de Velha Guarda* . Some arrangements of Pixinguinha from the series *Brasília Orquestra*[7]. and others, have been performed in two CDs directed by *Henrique Cazes* and released by the record labels *Kuarup* and *Biscoito Fino*.

4. CONCLUSION

By analysing the number of arrangements and original compositions presented by the orchestra of the radio program *O Pessoal da Velha Guarda*, the musicological universe of Pixinguinha can be studied under the orchestration said "Brazilian".

From new and old recordings and publications (music scores and books) dedicated to musical practises of the late 19th century and early 20th century and edited since the 1980s in Brazil new paths opening for musicians and researchers practise and value such musical experiences as well as know the works of numerous composers representing the music of that period.

5. REFERENCES

Aragão, Paulo (2001). *Pixinguinha e a gênese do arranjo musical brasileiro (1929 a 1935).* Master's thesis text. The graduate program in Music. Centro de Letras e Artes. Universidade do Rio de Janeiro.

Aragão, Pedro (2013). *O Baú do Animal:* Alexandre Gonçalves Pinto e O choro. Ed. Folha Seca. Rio de Janeiro.

Barata, Denise (2012). Permanências e deslocamentos das tradições musicais africanas na cidade do Rio de Janeiro. In *XXII Encontro da Associação Nacional de Pesquisa e Pós-Graduação em Música, ANPPOM,* João Pessoa.

Cabral, S. (2007). Pixinguinha: vida e obra. Rio de Janeiro, Fundação Nacional de Artes, FUNARTE.

Corrêa do Lago, Manoel, A. (org.) (2012). *O Boi no Telhado: Darius Milhaud e a música brasileira do modernismo francês.* São Paulo, IMS - Instituto Moreira Salles.

Fontenele, A.L. (2014). Pixinguinha em plena maturidade: os arranjos para um "Orquestra Brasileira". In: *Anais do VI Encontro de Musicologia de Ribeirão Preto.* USP.

Machado, Cacá (2007). *O enigma do homem célebre:* ambição e vocação de Ernesto Nazareth. Instituto Moreira Salles. São Paulo, 2007.

Paes, Anna. (2012). *Almirante e o Pessoal da Velha Guarda: Memória, História e Identidade.* Master's thesis text. The graduate program in Music. Centro de Letras e Artes da UNIRIO. Rio de Janeiro, 2012.

Paes Leme, B. (org) (2010). Pixinguinha Na Pauta: *36 arranjos para o programa O Pessoal da Velha Guarda.* Instituto Moreira Salles e Imprensa Oficial (SP).

_____ (2014), Pixinguinha – *Outras Pautas: 44 arranjos para o programa O Pessoal da Velha Guarda.* São Paulo: Instituto Moreira Salles/Imprensa oficial do Estado de São Paulo e Edições SESC - São Paulo.

Pinto, Alexandre, G (2009). *O Choro.* Rio de Janeiro, Fundação Nacional de Artes, FUNARTE.

Piedade, A. (2007). Expressão e sentido na música brasileira: retórica e análise musical. In *Revista Eletrônica de Musicologia,* Vol. XI.

Sandroni, C. (2012). *Feitiço Descente: transformações do samba no Rio de Janeiro (1917-1933).* 2ªed. Rio de janeiro. Zahar.

Severiano, Jairo (2008). *Uma história da música popular brasileira:* das origens à modernidade. Rio de Janeiro, Editora 34.

Tinhorão, José, R. (1991). *Os Sons do Brasil:* trajetória da Música Instrumental. São Paulo. SESC -Serviço Social do Comércio.

_____ (1997). *Música Popular Brasileira*: um tema em debate. (3ª edição revisada e ampliada), São Paulo, Editora 34.

_____ (2010). *História Social da Música Popular Brasileira.* (2ª ed.) São Paulo, Editora 34.

Vianna, Hermano (2007). *O mistério do samba.* (6ª ed.), Rio de Janeiro, Jorge Zahar/Editora. Ed. UFRJ.

Webern, Anton. (1984). *O Caminho para a música nova.* São Paulo. Editora Novas Metas.

5.1 Long Play, Compact Disc e Web site

Carvalho, Rosa, G. de. *A discussão do conceito de identidade nos estudos culturais.* In. http://encipecom.metodista.br/mediawiki/images/a/a2/GT3-_26_-_Identidade_conceito_celacom.pdf Acessed in: 13.04.2014. Universidade Católica do Rio Grande do Sul – PUCRS.

Cazes, Henrique (Prod) (1988). *Orquestra Brasília: o maior legado escrito de Pixinguinha.* Kuarup Discos, Rio de Janeiro.

_____ (2005), *Orquestra Pixinguinha.* Biscoito Fino. Rio de Janeiro.

Piedade, A. (2013). A teoria das tópicas e a musicalidade brasileira: relfexões sobre a retoricidade na música. In *El oído pensante 1(1),* http://ppct.caicyt.gov.ar/index.php/oidopensante, Acessed in: 23.08.2014.

Pixinguinha, A.R.V. (1972), *Pixinguinha e sua Gente – 1972 – Em tempo de Velha Guarda.* Fontana (Companhia Brasileira de Discos), Rio de Janeiro.

Prata, Sérgio. *Jacob do Bandolim*: breves anotações. Em: www.jacobdobandolim.com.br/jacob/biografia.php. Acessed in 17.03.2015.

Trotta, Felipe (2008). Samba Instrumental. In. *Músicos do Brasil: uma enciclopédia.* Em: www.musicosdobrasil.com.br. Acessed in 23.03.2015.

6. GLOSSARY

Batucada: Afro-Brazilian song & dance performed since the 17th century and consisting of sung verses (with choral response) accompanied de percussion instruments. Since the 1930s, the term signifies a specific type of samba with strong percussive accompaniment.

Batuque: the rhythm of batucada.

Choro: instrumental music of 19th-century origin, noted for virtuosity, improvisation, and counterpoint.

Malandro: bohemian figure of questionable character but great charm.

Maxixe: first genuinely Brazilian dance; a fusion of tango, habanera, and polka.

Modinha: old-fashioned lyrical, sentimental song of Portuguese derivation.

Roda de samba: samba circle; an informal gathering where samba is communally sung and played; also known as pagode.

Samba de roda: samba that is played in the state of Bahia.

A SIMPLE METHOD FOR MELODIC CLASSIFICATION BASED ON SCALE ANALYSIS

Anas Ghrab
Higher School for Music - University of Sousse (Tunisia)
anas@ghrab.tn

1. INTRODUCTION

One objective of the analysis of modal music is the comparison of melodies, which implies their classification. While conventional methods of analysis are used to enter a modal specificities bounded together, and the number is typically reduced, it is clear that the pertinence of the results is closely related to the amount of data analyzed. We propose a general method, implemented in Python, for the analysis of the scale of melodies and to measure the melodic nearby. We apply it to the analysis of a segmented file and several songs of women from different regions of Tunisia.

For the purpose of this analysis, we have implemented different functions in the python module **Diastema** [1] .

2. BACKGROUND

Several works the MIR field have focused on melodic analysis. Part of these works deal with melodies using a symbolic representation (De León et al., 2004; León & Iñesta, 2004; Li & Sleep, 2004; Frieler & Müllensiefen, 2005).

Other studies, such those of Bars Bozkurt (Bozkurt, Bozkurt) focused mainly on makamique analysis. As the special feature of *maqm* is in its intervallic system, which is not taken into consideration by the symbolic data, Bozkurt use different basic detection algorithms to obtain representative frequencies of the melody. These frequencies, retuned into the same octave and converted using Holderian Comma, allowed the representation of histograms of main notes of melodies (template-matching). Ioannidis et al. (2011) present an extension to work Bozkurt using HPCP. The willingness of these approaches is to determine the scales of the main modes of Turkish music or to identify the *maqm* of a music compared to a predetermined scale (theoretical tuning systems). We can consider the we are here in a supervised approach.

The method presented here is different. Our main goal is to analyze and classify melodies in an unsupervised manner. This approach should allow us to address and redefine *makam*-s and modal scales of different art music in a broader context. In addition, it is applicable to non-scholarly traditional music, where melodies do not belong to a pre-established modal theory.

[1] https://github.com/AnasGhrab/diastema

3. GENERAL METHODOLOGY

This method consists of an extraction of the fundamental frequencies of each melody, where we use the Predominant Melody Extraction algorithm (Salamon & Gómez, 2012) which is relatively robust for the detection of the melody in a complex context.

A Kernel Density Estimation is then applied to the frequency list to obtain Probability Density Functions (PDF-s) related to dominant frequencies in the melody. The peaks of the PDF gives as the frequencies that represent the scale, which are faced to most known intervals in an epimoric (n+1/n) definition. For that we need first to detect the frequency that could represent the tonic note. We detail below our approch for that.

When comparing melodies from different files, we have to transpose them on a single reference frequency. Then we classify Probability Density Functions by the linear Correlation Coefficient : the classification of PDFs should match a melodic nearby.

4. THE DIASTEMA TOOL AND EXAMPLES

4.1 First example

Our first example is the analysis of Violin Taksim.

4.1.1 PDF-s of the different segments

4.1.2 The detection of the tonic frequency-note

Useing different percentages of the last frequencies :

```
        [1.5%,2%,5%,10%,15%]
P1   :  [318, 318, 318, 318, 318]
P2   :  [318, 318, 318, 318, 347]
P3   :  [318, 343, 318, 318, 318]
P4   :  [316, 316, 316, 318, 318]
P5   :  [318, 318, 318, 318, 318]
```

4.1.3 The global PDF and the scale

```
[['0.00', '1/1', '+', '0.00'],
 ['36.54', '12/11', '-', '1.25'],
 ['88.39', '9/8*12/11', '-', '0.56'],
 ['114.26', '4/3', '-', '10.68'],
 ['146.52', '4/3', '+', '21.58'],
 ['181.95', '3/2', '+', '5.86'],
 ['209.70', '3/2*10/9', '-', '12.15'],
 ['258.89', '3/2*6/5', '+', '3.62']]
```

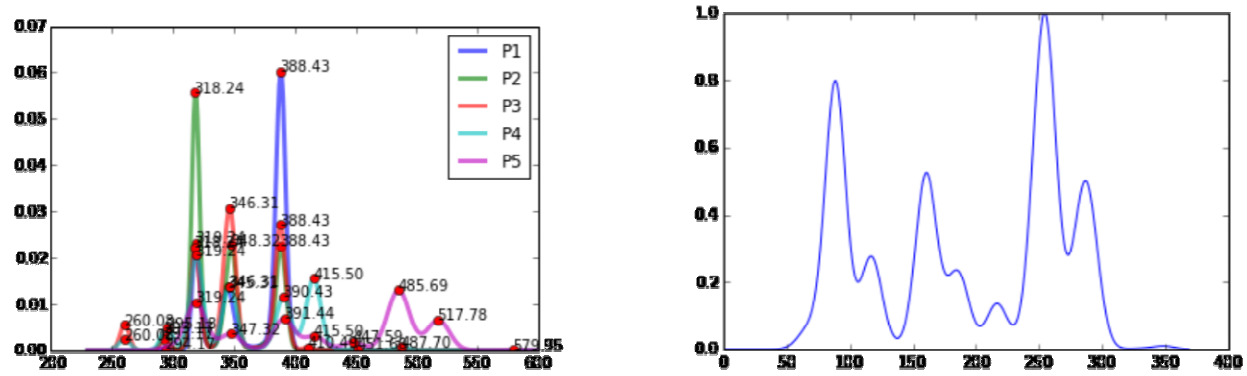


Figure 1: PDF-s of the different segments; Global PDF from all segments

4.2 Second example

303 audio files of traditionnal women songs from different regions of Tunisia. They are field recordings made between 2007 and 2015 [2] .

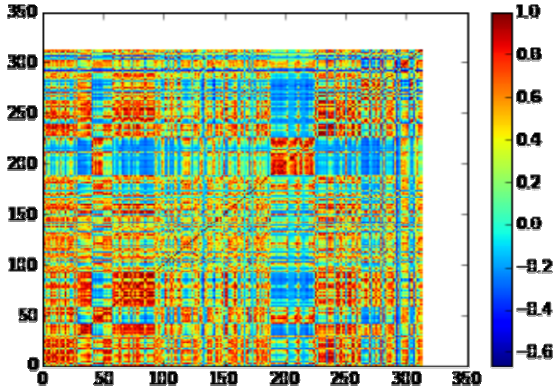


Figure 2: Self-similarity matrix

5. CONCLUSION

This paper presents a musicologial point of view. Its approach has to be extended with specialists in data-mining and programmers.

6. REFERENCES

Bozkurt, B. An automatic pitch analysis method for turkish maqam music. *37*(1), 1–13.

De León, P. J. P., Pérez-Sancho, C., & Inesta, J. M. (2004). A shallow description framework for musical style recognition. In *Structural, Syntactic, and Statistical Pattern Recognition* (pp. 876–884). Springer.

Frieler, K. & Müllensiefen, D. (2005). The simile algorithm for melodic similarity. *Proceedings of the Annual Music Information Retrieval Evaluation exchange*.

[2] See the collection "Chants de femme dans les regions tunisiennes" at the catalogue of the National Sound Archives (http://phonotheque.cmam.tn), hosted by the Centre of Arab and Mediterranean Music (http://www.cmam.tn).

Ioannidis, L., Gómez, E., & Herrera, P. (2011). Tonal-based retrieval of Arabic and Middle-East music by automatic makam description.

León, P. J. P. D. & Iñesta, J. M. (2004). Statistical description models for melody analysis and characterization. In *In Proceedings of the 2004 International Computer Music Conference*, (pp. 149–156).

Li, M. & Sleep, R. (2004). Melody classification using a similarity metric based on kolmogorov complexity. *Sound and Music Computing*, 126–129.

Salamon, J. & Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. *Audio, Speech, and Language Processing, IEEE Transactions on*, *20*(6), 1759–1770.

TRANSFORMATION OF A BERTSO MELODY WITH COHERENCE

Izaro Goienetxea[1] **Darrell Conklin**[1,2]
[1] Department of Computer Science and Artificial Intelligence
University of the Basque Country UPV/EHU, San Sebastián, Spain
[2] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
{izaro.goienetxea, darrell.conklin} @ehu.eus

1. INTRODUCTION

The topic of automatic generation of music has existed for a long time, and many different approaches have been developed. One of these approaches is the use of statistical models. Statistical models of symbolic representations of music have been used in many works in the computational modeling of different music styles, like folk music. These models are able to capture some musical features, making it possible to generate new musical sequences that reflect an explicit musical style (Conklin, 2003; Dubnov et al., 2003).

Statistical models can also be used to transform pieces, by extracting the structure of a piece, describing it using semiotic labels, and sampling new pieces conserving that structure. This way a musical cohesion can be given to new productions (Conklin, 2003).

Bertsolaritza or bertsolarism is the art of singing improvised songs in Basque (bertsos), respecting various melodic and rhyming patterns. There is evidence of bertso singing and written bertso poem samples since the 15th century, and it is a very popular art nowadays in the Basque Country. Bertsos are sung in many different occasions, like informal lunches with friends, homage ceremonies or competitions and any topic can occur in a bertso. Many bertsolari competitions take place every year in the Basque Country, and every four years the national championship final is held, with around 15000 people in attendance.

In the work described herein a bertso melody is transformed to create new melodies, maintaining its original coherence structure.

2. METHODS

2.1 Bertso

Bertsolaritza is defined as a sung, rhymed and metered discourse by the book The art of bertsolaritza: improvised Basque verse singing, written by Garzia et al. (2001).

Bertso Doinutegia [1] is a collection of 3059 bertso melodies, created by Joanito Oiartzabal and published for the first time on 1995. It is updated every year by Xenpelar Dokumentazio Zentroa [2] with new melodies that are used on competitions or bertso exhibitions. Entries in the collection have a melody name, the name or type of the strophe, type of the melody (genre), creator, bertsolari who has used it, name and location of the person who has collected the melody and year of the collection. Melodies are classified into 17 different types or genres, and this classification is done based only on the melody. Some of the melodies in the collection have links to recordings of exhibitions or competitions where those melodies were used.

[1] http://http://bdb.bertsozale.eus/en/web/doinutegia/bilaketa
[2] http://bdb.bertsozale.eus/en/info/7-xenpelar-dokumentazio-zentroa

2.2 Semiotic Structure

Music structure description is currently a scientific challenge, and it can be approached in several ways. In this work semiotic structure is used to describe the structure of a bertso melody. Semiotic structure represents the similarities and internal relations of structural segments, using labels to identify similar segments (Bimbot et al., 2012). In the work described herein, the semiotic structure of a bertso melody is described by assigning a different semiotic label to each different note. For transformation, this label structure is taken as a constraint, specifying which notes must have the same pitch value.

Taking the whole structure as a constraint can be too restrictive when generating new melodies, since it does not allow a wide variety of different transformation. A melodic reduction strategy has been applied, to put the labels only on the more significant notes, allowing any pitch on the less important positions. These positions are labeled with a label X, where X matches any note. This method identifies the notes where the melodic direction changes, similar to what the contour reduction viewpoint of Conklin & Anagnostopoulou (2006), does but selecting the note previous to the position where the new contour is found, and puts label X on the other notes.

On the piece that is transformed in this work 42 Xs have been set out of 100 notes.

2.3 Sampling

Since in bertsos meter is a very important feature, in this work the rhythmic structure of the original piece is conserved, and a new melody line is created. A statistical model has been built, computing the probabilities of transitions between pitch contour values of 15 bertso melodies of the corpus described in Section 2.1. To do so, a five point contour (leap down, step down, repetition, step up, leap up) viewpoint has been computed, as in Conklin & Bergeron (2008), steps involving a motion of one or two

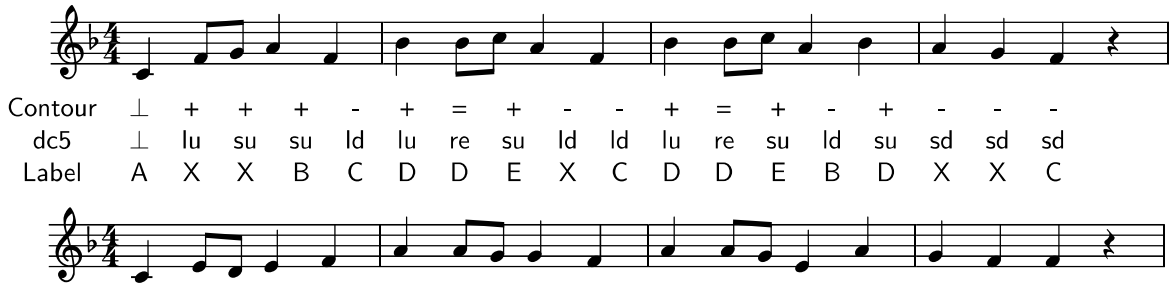


Figure 1: Segment of a transformation (bottom) of a bertso melody (top) with melodic contour, five point contour (dc5) and label of each note, determined by contour reduction.

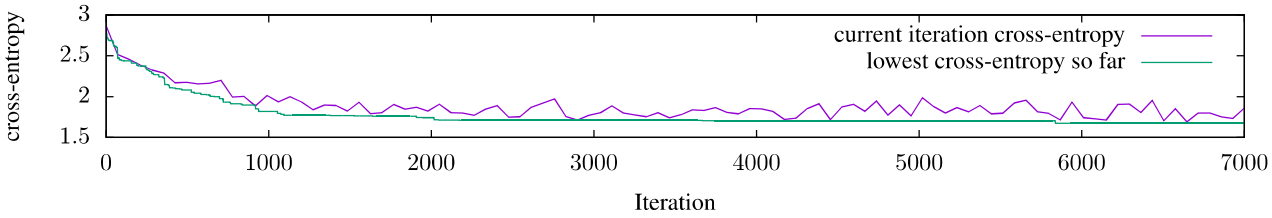


Figure 2: Cross-entropy of the piece through the iterations and the lowest cross-entropy reached.

semitones, and leaps a motion greater than two semitones.

A *stochastic hill climbing* method has been used for sampling, and this process has been iterated 10^4 times. A random piece with the same semiotic structure as the original song is taken as a starting point, and in each iteration a random location i in the piece is chosen. A random element $e_i \in \xi$, where ξ is an event space which describes the set of possible music notes, is substituted into that position. In the bertso melody transformed in this work a 9 element vocabulary ξ is used, with pitch numbers from 60 to 74. If the semiotic label that corresponds to position i is not X, the substitution must be done in all the positions in the piece that have the same semiotic label, producing a new piece $\mathbf{e} = e_1, ..., e_i, ...e_l$ with an updated probability. The probability of the piece is computed using the single viewpoint model described in Conklin (2013) and presented in the equation below. Letting $v_i = \tau(e_i|e_{i-1})$ be the contour feature of event e_i in the context of its preceding event e_{i-1}, the probability of the piece e is computed as:

$$P(\mathbf{e}) = \prod_{i=1}^{\ell} P(v_i) \times P(e_i|v_i, e_{i-1})$$

$$P(e_i|v_i, e_{i-1}) = |\{x : \tau(x|e_{i-1}) = v_i\}|^{-1}$$

The probability is used to measure the *cross-entropy* of the piece; the mean negative log probability of an event in the piece, defined by $-\log_2 P(\mathbf{e})/\ell$. If the new cross-entropy is lower than the last saved one, it is saved and next iteration is executed on the new piece.

3. RESULTS

A small section of a bertso melody transformation is shown on Figure 1. The contour sequence of the original piece is shown, as well as its five point contour (dc5) sequence, where ld represents a leap down, sd is a step down, re is a repetition, su is a step up and lu is a leap up. The semiotic structure is represented with labels from A to E. The graph on Figure 2 shows how the cross-entropy of the piece varies through the iterations, in purple. The line has been smoothed with splines, so the direction of the cross-entropy can be seen.The lowest cross-entropy is reached before iteration 6000. Although the stochastic hill climbing method does not guarantee to find the optimal piece, it does improve the initial piece as seen on the green line of Figure 2.

4. CONCLUSIONS

In this work we present a method to transform a bertso melody preserving its musical structure, described by its semiotic structure. Semiotic structure is a good way to describe the coherence of the original piece, since all the repetitions on the piece that is being transformed are captured, and are going to be present on the transformation. Since assigning labels to all the notes on a piece can be too restrictive on sampling and it does not allow a wide variety of transformations, a contour reduction algorithm is used to allow some of the notes on the piece not to have a constraint. Using this method many different resulting melodies have been obtained, keeping the structure of the original piece, but they do not necessarily keep its repetitions. This can be a problem for bertso melodies, since they should be simple in order to make the process of creating and singing the bertso easier to the bertsolari. To solve this issue, pattern discovery methods in conjunction with semiotic structure labels are being explored.

5. ACKNOWLEDGEMENTS

This research is supported by the project Lrn2Cre8 which is funded by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 610859.

6. REFERENCES

Bimbot, F., Deruty, E., Sargent, G., & Vincent, E. (2012). Semiotic structure labeling of music pieces: Concepts, methods and annotation conventions. In *13th International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 235–240)., Porto, Portugal.

Conklin, D. (2003). Music generation from statistical models. In *Proceedings of the AISB 2003 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences*, (pp. 30–35).

Conklin, D. (2013). Multiple viewpoint systems for music classification. *Journal of New Music Research*, *42*(1), 19–26.

Conklin, D. & Anagnostopoulou, C. (2006). Segmental Pattern Discovery in Music. *Informs Journal on Computing*, *18*, 285–293.

Conklin, D. & Bergeron, M. (2008). Feature set patterns in music. *Comput. Music J.*, *32*(1), 60–70.

Dubnov, S., Assayag, G., Lartillot, O., & Bejerano, G. (2003). Using machine-learning methods for musical style modeling. *Computer*, *36*(10), 73–80.

Garzia, J., Egaña, A., & Sarasua, J. (2001). *The art of bertsolaritza: improvised Basque verse singing*. Donostia, Bertsozale Elkartea; Andoain, Bertsolari Liburuak, 2001.

MELODIC KEY PHRASES IN TRADITIONAL CRETAN DANCE TUNES

Andre Holzapfel
Boğaziçi University & ITÜ MIAM
{andre}@rhythmos.org

1. INTRODUCTION

In the Mediterranean island of Crete, local music traditions were a rather rural phenomenon about 40 years ago. In the urban environments of Crete, European or Greek art musics were preferred, and little value was attached to the local music idiom, which is referred to as Cretan music (*kritiká*) by the the local population. Even in village festivities, Cretan music represented only a part of the repertoire, while "European" dances (*evropaiká*) like the Waltz and the Tango played a very important role in the rural social events. At some point in the 1980's, however, the first concert events were established in the rural environments that featured concerts of Cretan music. Apparently, this shift of the performance context of Cretan music from local festivities to the stages of professionally programmed concert events marked a turning point towards an increasing popularity of Cretan music in all parts of Crete, including the urban environments. Nowadays, we encounter Cretan music all over the island of Crete. During the summer months, the municipalities of all cities and village areas organise regular concert events that feature Cretan musics as well as musics from other parts of Greece and beyond. In the year 2011 I counted a total of five radio stations operating in the city of Heraklion, with a total of 200.000 inhabitants, that broadcast exclusively Cretan music. A large number of dance schools in all areas offer lessons on Greek folk dances, and the curriculum of every such school includes Cretan dances, with the number of different Cretan dances being taught nowadays reaching the number of 20.

Cretan dances share a large set of characteristics. Most dances are circular dances, performed counterclockwise. Specific emphasis is given to the movements of the feet, while the upper part of the body remains widely motionless. The dancer leading the circle has, at least for many of the dances, the freedom to improvise on the standard sequence of steps. Regarding their musical properties, all Cretan dances are usually notated as having $2/4$ meters. Rhythmic accompaniment, in most cases provided by one or more Cretan lutes, makes use of a limited set of rhythmic patterns. Especially for the faster leaping dances, which make up the largest group in number, these patterns, depicted in Figure 1, are shared between the different dances.

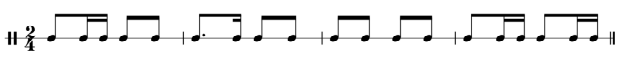


Figure 1: Rhythmic patterns reported by Kaloyanides (1975) for Cretan leaping dances

In terms of melodic content, the Cretan dance tunes are built from the concatenation of short melodic phrases, referred to as *kontiliés* by the local musicians, and played in most cases by the Cretan fiddle (*lýra*) or violin. These melodic phrases span a duration of two bars for most of the leaping dances, and many of these phrases are related to certain musicians of previous generations, and are often assigned to a specific dance. The cataloguing of melodic phrases in Greek music was approached, for instance, by Theodosopoulou (2004) for specific Cretan tunes, and by Sarris (2007) for bagpipe tunes from the Greek mainland. Locating important melodic patterns has been previously attempted by Conklin & Anagnostopoulou (2011) using existing transcriptions for analysis. However, the resulting patterns were usually shorter than the two- or four-bar length of the Cretan *kontiliés* and have a less obvious musical meaning as identifier of the specific tune. The mentioned attempts make use of manual transcription and while they provide valuable insight into the sequential structure in which *kontiliés* are applied in various tunes, to the best of my knowledge so far no analysis has been published that aims at the identification of meaningful key phrases. Such key phrases might serve the function to identify certain dances, and help the participants in a music event to differentiate between dance tunes. In this paper, a method that we previously presented for the alignment between performance recordings and notations of Turkish Makam music (Şentürk et al., 2014) will be applied to analyse the melodic phrases contained in a set of leaping dance recordings. This analysis works automatically to a large extent, and therefore enables for an analysis of a larger corpus of recordings.

In this paper, Section 2 will give a short summary of the computational methodology, referring to the initial papers for further detail in order to restrain the technical description within this paper. In Section 3, a set of performances of five leaping dances is analysed and the question is addressed in how far these dances can be differentiated by using melodic information. Since the results demonstrate the importance of melodic phrases as identifiers for the dances, I do a first steps towards a documentation of the melodic key phrases in Section 4, and Section 5 will conclude the paper.

2. COMPUTATIONAL METHOD

In performances of Cretan music, the most common lineup consists of Cretan *lýra* or violin as main melodic instrument, and one ore more accompanying Cretan lutes. If

more than one lute is present, usually one lute doubles the main melody with the lead instrument in a heterophonic way. In order to estimate the melody in such (non-monophonic) recordings, I applied the analysis algorithm presented by Salamon & Gómez (2012), which is tailored for the task of estimating the lead melody out of a music signal. Using these pitch estimations as input, I applied the algorithm that we proposed for the scope of matching two melodies (Şentürk et al., 2014). This algorithm takes two vectors that contain pitch values, and tries to find (partial) occurrences of the melody of the first vector in the melody described by the second vector. I will refer to the first vector as the query and the second vector as the target melody. For each performance recording to be analysed I annotated beats and downbeats using the approach presented in Holzapfel et al. (2014). Melodic patterns in Cretan music typically have a length of either two or four measures in the analysed dances. Using the obtained beat and downbeat annotations, the recordings can be segmented into slices of four measure length. The pitch estimations obtained from these slices will be used as queries, trying to locate their re-occurrences in other recordings.

The method of pattern matching (Şentürk et al., 2014) computes differences in pitch values, but allows for a certain amount of inaccuracy for the matching. To be precise, we demanded the matched melodic patterns to be at least of one measure length, and that at most a duration related to an eighth-note can be different between two matched patterns. The pitch tolerance for two pitches to be judged as identical is a quarter tone. These tolerances make the method robust to technical problems in the pitch estimation, and to slight differences between the performances caused, for instance, by ornamentations.

3. PHRASE DIVERSITY IN LEAPING DANCES

The music corpus used in this paper contains 22 audio recordings, each assigned to one of five different leaping dances. The recordings are obtained from commercially available recordings of renowned Cretan musicians. Currently, a larger set of recordings is prepared for an analysis, including material from our field work in Crete. The analysis of this larger corpus will be subject of a future publication, and will enable to obtain a more detailed understanding of melodic phrases than the insights obtained from the relatively small corpus in this paper. Table 1 specifies the names of the dances as they are usually referred to in Crete, along with the number of pieces and bars of $2/4$-meter for each of the dances.

I applied the approach described in the previous section, and counted the number of times in which a query phrase from a specific dance recording is encountered in a different performance. I categorised each of these encounters according to the class of dance, in which the matched target phrase was found. This procedure provides us with a set of five histograms, one for each dance, as depicted in Figure 2. In the histogram for Dance A, for instance, the five bars denote the probability to encounter a phrase from Dance A in a different recording of Dance A, B, C, D or E, respectively. The histograms were sorted according to the clarity of results; Especially the histograms for the *Anogianós* and *Ethianós* dance tunes indicate that these dances have a melodic material that is not shared with the other dances (very small bars for dances different from the query melody). However, even for Dance E, melodic phrases reoccur most of the time in dance tunes of the same dance (the fifth bar marked 'Sit' is the largest).

Dance	Number of Pieces	Number of Bars
Anogianos	4	819
Ethianos	2	396
Maleviziotis	4	953
Sousta	8	1873
Sitiakos	4	1278
Total	22	5319

Table 1: Number of samples per dance and number of bars available for the analysis in the following experiments.

The clarity of the differentiation especially for dances A and B could be presented as a result of geography; Both *Anogianós* and *Ethianós* are dances from remote mountain areas of Crete. On the other hand, the recent historical development of performance context seems to be an even more likely explanation to me; Both dances do until now not play a major role in the concert performances, but appear only if demanded by the audience. On the other hand, *Soústa* and especially *Maliviziotis* are integral part of most festivities in Crete, and therefore represent dances that underlie a lively process of permanent re-invention at every moment of performance. This is caused by the fact that the sequence of melodic phrases used in each performance is improvised to a wide extent and can be changed according to the demands of the performance situation. This dynamic process might have lead to a stronger exchange of melodic phrases with other dances. A particular role is taken by the *Sitiakós* dance, which is considered a variant of the *Malevizió́tis* by many informants from central Crete, something that is rejected by informants from the Eastern part of Crete, where this dance has its origins. In its histogram (Dance E) its strong relation with the *Malevizió́tis* can be recognized (the third bar is the second highest after the bar relating the dance to itself). However, an answer regarding origins shall not be attempted at this place. Rather than that, I would like to point out that these dances are rather an impressive documentation how local expressions travel and interact with other expressions, a process of constant change and negotiation that was also proposed as a redirection of research agenda in anthropology recently (Clifford, 1997).

Both the presented analysis and information from interviews support that the melodic phrases that occur only within a particular dance have some significance for the identity of the dance tune. Within an analysis that will be presented as an upcoming book chapter (Holzapfel, 2015), tempo aspects and aspects of rhythmic accentuation were examined as well. Tempo shapes a clear difference be-

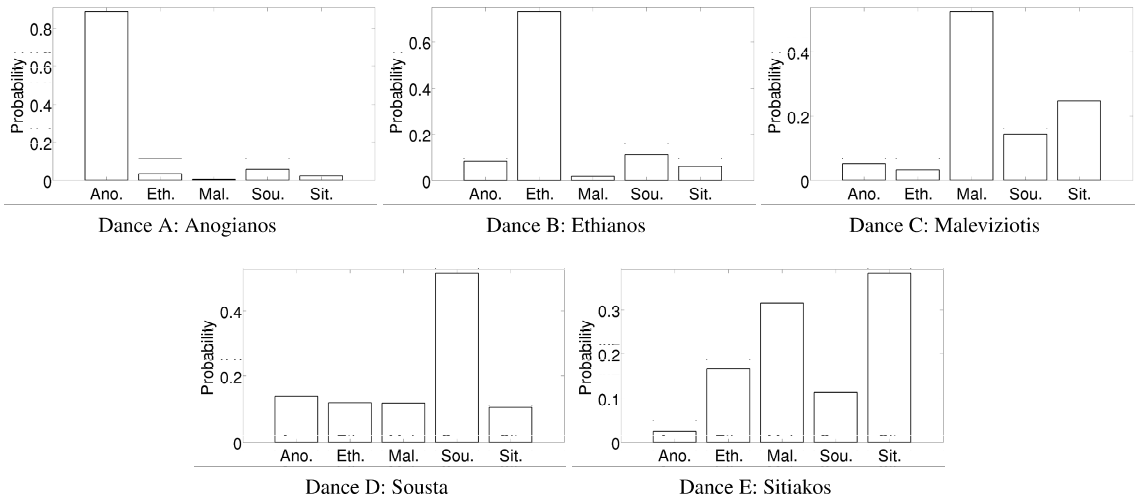


Figure 2: Distribution of the detected melodic patterns for the five dances.

tween some dances, such as the *Malevizιótis* and the *Soústa*, two dances that differ widely regarding the body motions and spatial organisation, with the former being a circle dance and the latter being a partner dance. The overall difference in physical expression of these dances is clearly related to their tempo differences, with the *Sousta* being one of the slowest leaping dances in Crete. Rhythmic accentuation, on the other hand, appears to play a more subtle role than tempo differences. While some lute players state that they place rhythmic accents differently depending on the dance, these differences depend on the players, and it will need to be established in future field work in how far different stroke patterns on the lute are used depending on the dance. The role of the rhythmic accentuation by the *lýra* was emphasised by several informants in Crete. For instance, the dance teacher Vangelis Stafilas (2011) stated that the bow stroke for the *Pentozális* dance would be very rough compared to the ones of *Malevizιótis* and especially *Soústa*. Within various *lýra* seminars, techniques to vary the bow strokes are taught (Spiridakis, 2013), but a detailed study of these techniques remains to be conducted.

However, my analysis as well as the conducted interviews assign a major importance to the choice of the *kontiliá* for the differentiation between dances. Dance teacher Yannis Panagiotakis (2013) has no doubt that the recognition of the dance tune is based on memorising the melodies of the *kontiliés* belonging to all the different dances. This is consistent with my observations during a listening test that I conducted with dancing teachers, experienced dancers, and Cretan citizens without larger dance experience (Holzapfel & Stylianou, 2011). I asked the subjects to rate the similarity between a query and two recordings taken from different performances than the query. One of these two recordings was from the same dance as the query (but from a different performance), while the other was from a different dance. I observed that the dance teachers did their rating after listening to about a second of each recording, meaning that they did not even listen to the whole phrases, but the first notes were sufficient for them to relate the sound to a memorised phrase. The experienced dancer usually listened to the whole recording (of about 10s length), while the inexperienced often chose to listen to the samples several times looking for the important cues to guide their decisions. While these results are rather informal due to the small group of subjects, they confirm the statement of Yannis Panagiotakis, and the results depicted in this section that state that the melodic content of the short phrases is a major cue for identifying the dances.

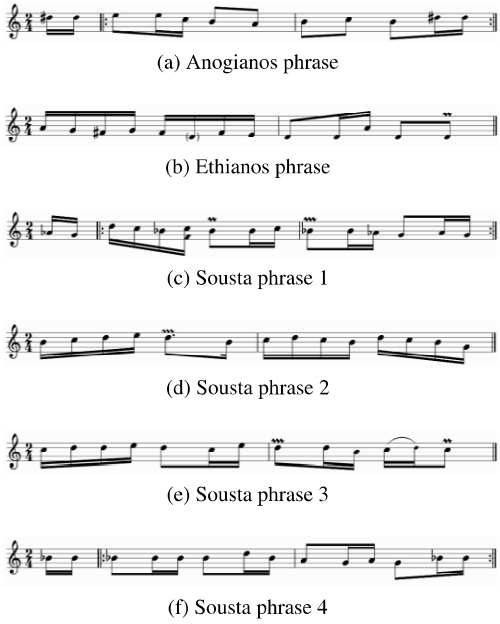
4. A CATALOGUE OF KEY PHRASES

In Figure 3 examples for *kontiliés* are depicted that appeared frequently, but almost exclusively within a specific dance within the corpus. Most patterns found so far are clear markers for the dances they were found in, judging by the expertise of my collaborators and me. It is worth pointing out again that the identification of these phrases works almost completely automatically, with the only manual steps being the selection of recordings and an inspection of the automatic beat annotations.

Only throughout the recent years, Cretan music is taught to a wider extent within music seminars in Crete to interested musicians. The first person to systematically approach teaching of Cretan music was the *lyráris* Kostas Mountakis, in 1979 and throughout the 1980's, interestingly in parallel to the appearance of the first Cretan music concert events. Another important catalyst for the increasing interest in Cretan music was the foundation of the *Labyrinth* music workshops [1], initiated by the musician Ross Daly, and since then supported by a large group of international music teachers. While the teaching within the Labyrinth workshops embraces many music cultures, it contributed to the interest of young musicians into studying Cretan music within the setup of a very focused seminar environment.

The demand for seminars and the increased interest in Cretan music on concert stages interacts with the increas-

[1] www.labyrinthmusic.gr



(a) Anogianos phrase

(b) Ethianos phrase

(c) Sousta phrase 1

(d) Sousta phrase 2

(e) Sousta phrase 3

(f) Sousta phrase 4

Figure 3: Some examples of phrases that were found to appear frequently exclusively in one particular dance.

ing number of tunes that are discovered and re-invented in various areas in Crete. This process is driven by groups of local researchers, with and without academic affiliations, and feeds the increased interest of the public into the tunes that identify various local micro-cultures in Crete. An increasing number of melodic phrases is the outcome, and discovering these phrases and reflecting on their meaning in collaboration with the local communities is a driving motivation for my field work, as I will point out in the concluding remarks.

5. CONCLUSION

It is apparent from my experiments as well as from information from various interviews and field observations that melodic phrases of the kind depicted in Figure 3 carry a diverse set of meanings for the local musicians. These meanings go beyond the identification of a dance tune. These meanings have historical and aesthetic dimensions, and by bringing the discovered analysis results back into the field, I hope to obtain deeper insights into these interpretations. The melodic key phrases presented in this paper will be discussed with musicians in Crete, during my subsequent field work in Crete during summer 2015. Apart from verifying their ability to identify dances, the goal is to discover the context and the genesis of these patterns, and to reveal the different meanings that they might have to individual musicians. Such kind of a discussion regarding the inherent structure of a music is hard to accomplish using purely ethnographic approaches, and the combination of ethnography with the presented analytic results can hopefully provide intuitive ways to establish a deeper discourse with the local musicians. Apart from that, as mentioned above, I hope that such a research can support the ambitious teaching activities of local musicians. Such a support can be provided, for instance, by providing printed material including transcriptions of key phrases. While such transcriptions might not be of interest for most experienced Cretan musicians, especially musicians coming from written music traditions can be supported in obtaining an understanding of Cretan music more easily. Since many musicians embrace the usage of technology in various ways, I consider this an offer I can make to the community.

6. ACKNOWLEDGEMENT

I would like to thank Michael Hagleitner for the great research collaboration over the years. I would also like to thank my advisor Robert Reigle for his continuing support. This work is supported by a Marie Curie Intra-European Fellowship (grant number 328379).

7. REFERENCES

Clifford, J. (1997). *Routes: Travel and Translation in the Late Twentieth Century*. Cambridge, MA: Harvard University Press.

Conklin, D. & Anagnostopoulou, C. (2011). Comparative pattern analysis of cretan folk songs. *Journal of New Music Research*, *40*(2), 119–125.

Şentürk, S., Holzapfel, A., & Serra, X. (2014). Linking scores and audio recordings in makam music of Turkey. *Journal for New Music Research*, *43*(1), 34–52.

Holzapfel, A. (2015). Patterns of identity: Rhythmic and melodic aspects of Cretan leaping dances. In *Music on Crete, Traditions of a Mediterranean island.* Vienna Series in Ethnomusicology, in press.

Holzapfel, A., Krebs, F., & Srinivasamurthy, A. (2014). Tracking the "odd": Meter inference in a culturally diverse music corpus. In *Proceedings of ISMIR - International Conference on Music Information Retrieval*, (pp. 425–430)., Taipei, Taiwan.

Holzapfel, A. & Stylianou, Y. (2011). Scale transform in rhythmic similarity of music. *IEEE Transactions on Audio, Speech and Language Processing*, *19*(1), 176–185.

Kaloyanides, M. (1975). *The Music of Cretan dances, a study of the musical structures of Cretan dance forms as performed in the Irakleion province of Crete*. PhD thesis, Wesleyan University.

Panagiotakis, Y. (2013). Personally conducted interview.

Salamon, J. & Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech and Language Processing*, *20*(6), 1759–1770.

Sarris, H. (2007). *The bagpipe in Evros (in Greek language)*. PhD thesis, University of Athens. in Greek language.

Spiridakis, Z. (2013). Music Seminar in Houdetsi.

Stafilas, V. (2011). Personally conducted interview.

Theodosopoulou, I. B. (2004). *Methodology of morphological analysis and analytic data of small rhythmic patterns of cretan folk music, (in Greek Language)*. Athens: Kultura.

THE DEEP HISTORY OF MUSIC PROJECT

Armand Leroi, Imperial College London
a.leroi@imperial.ac.uk
Matthias Mauch, Queen Mary University of London
m.mauch@qmul.ac.uk
Pat Savage, Tokyo University of the Arts
patsavagenz@gmail.com
Emmanouil Benetos, Queen Mary University, London
emmanouil.benetos@qmul.ac.uk
Juan Bello, New York University,
jb2843@nyu.edu
Maria Panteli, Queen Mary University, London
m.panteli@qmul.ac.uk
Joren Six, University of Ghent
joren.six@ugent.be
Tillman Weyde, City University London
T.E.Weyde@city.ac.uk

1. INTRODUCTION

Music, like language and genes, is the product of a descent–by–modification process (MacCallum et al., 2012). As such, the current distribution of music styles around the world should reflect the history of human migration and cultural diffusion (Lomax, 1968). However, where geneticists and linguists have developed sophisticated techniques for reconstructing that history, ethnomusicologists have largely abandoned large-scale comparative studies (Leroi & Swire, 2006; Savage & Brown, 2013). Here we outline a proposal to revive comparative musicology using recently digitized ethnomusicological archives and MIR technology. Our study has three objectives: (i) To determine global distribution of musical style; (ii) To investigate the relationship between patterns of musical, linguistic and genetic diversity; (iii) To construct a large open-access database containing MIR features and metadata from traditional music.

2. METHODS

2.1 Sources

Our study relies on recently digitized ethnomusicological archives (Cornelis et al., 2005; Weyde et al., 2014). We have access to the following archives (number of tracks, thousands): British Library, London (33k), Royal Central Africa Museum, Tervuren (30k), Smithsonian, Washington DC (30k), as well as some smaller archives. We are currently in discussions to obtain the holdings of the Centre de Recherche en Ethnomusicologie, Paris and Ethnologisches Museum, Dahlem, both of which contain tens of thousands of tracks. These tracks have been filtered for music that we believe was primarily composed for oral rather than mechanical transmission. We have also standardized the metadata from these diverse sources by means of a standardized geographic and cultural group ontology. In all, we estimate that our initial database will contain ≈75k useful recordings from $>$ 300 cultures.

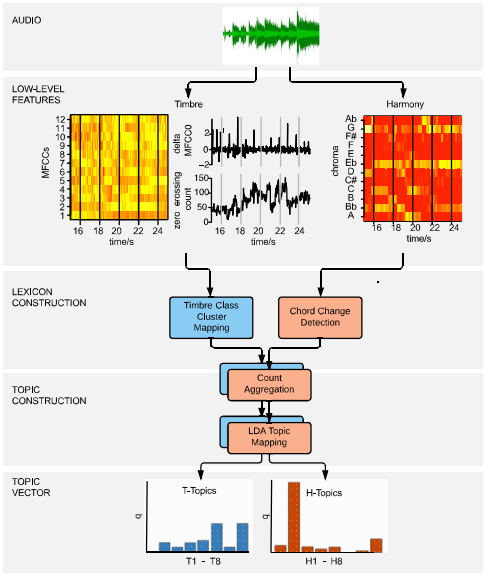


Figure 1: Data processing pipeline illustrated with a segment of Queen's *Bohemian Rhapsody*, 1975. From (Mauch et al., 2015).

2.2 MIR features

We have examined a variety MIR features, focusing on melodic, rhythmic and timbral descriptors that are not specific to Western-tradition music. To do this, we have tested existing descriptors, or modifications of them, against two sets of audio: a small set of synthesized audio designed to vary rhythm and melody systematically, and another set of cross-cultural real-world recordings (Panteli & Mauch, 2015). The rhythmic decriptors that we have tested are Onset Patterns, Fluctuation Patterns and Scale Transform; the melodic descriptors are Pitch Bihistograms, Magnitudes of the 2D Fourier Transform and Intervalgrams. For rhythm, the best performing descriptor was a modification of Onset Patterns, and for melody, the best performing descriptor was a modification of Intervalgrams.

These features were then processed further using a technique inspired by text-mining that we have successfully used in a large study of American popular music (Mauch et al., 2015), shown in Figure 1. Briefly, the features were

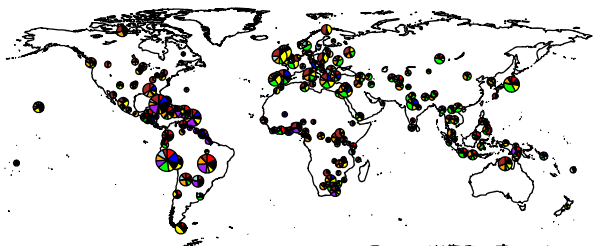


Figure 2: Global map of 7 musical styles.

discretised into "words" resulting in a rhythmic lexicon (R-lexicon), a timbral lexicon (T-lexicon) and a melodic lexicon (M-lexicon). These features were then combined into combinations of musical "words", or "Topics". Each song, then, is described by a vector of 10 R-, T- or M- Topics, making a total of 30 higher-level features.

3. RESULTS

In order to determine whether our features have any power to uncover structure in the world's music, we have been studying a subset of the Smithsonian database. Since most of the variation in music is located within, rather than among, cultures (Savage & Brown, 2014), we think that the basic unit of analysis should be Styles (c.f. Mauch et al. (2015)). To take a first look at such Styles we have carried out K-means clustering on our Topics and mapped their geographic distribution (Figure 2). These results are encouraging for they suggest that particular Styles are indeed enriched in certain parts of the world and hence that our data do capture at least some global musical structure.

4. DISCUSSION

We have only begun to analyse our data. Much remains to be done in terms of filtering our songs further and refining the basic features, Topic analysis, and clustering procedures. Once we have done that, we will proceed to examine the distribution of stylistic patterns formally via Bayesian spatial models in order to distinguish stylistic similarities due to diffusion from those that are due to convergent evolution. The resulting spatial analyses will then be combined with geographic genetic and linguistic data (e.g., Leslie et al. (2015)) in order to test causal, historical, explanations for the distribution of musical Styles.

Although we will initially focus on continent-scale musical diversity, our techniques and data can be used for studies any scale. We envision that our data will form the basis for a publicly accessible database of the world's music that will expand as new MIR features are developed and additional archives are digitized. To that end, we welcome collaboration from MIR experts, ethnomusicologists and archive-curators.

5. 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). *IASA JOURNAL*, (26), 35–43.

Leroi, A. M. & Swire, J. (2006). The recovery of the past. *World of Music*, *48*.

Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E. C., Cunliffe, B., Lawson, D. J., Falush, D., Freeman, C., Pirinen, M., Myers, S., Robinson, M., Donnelly, P., Bodmer, W., Control, W. T. C., & Genetics, I. M. S. (2015). The fine-scale genetic structure of the British population. *Nature*, *519*(7543), 309+.

Lomax, A. (1968). *Folk song style and culture*. Washington, D. C.: American Association for the Advancement of Science.

MacCallum, R. M., Mauch, M., Burt, A., & Leroi, A. M. (2012). Evolution of music by public choice. *Proceedings of the National Academy of Sciences*, *109*(30), 12081–12086.

Mauch, M., MacCallum, R. M., Levy, M., & Leroi, A. M. (2015). The evolution of popular music: USA 1960–2010. *Royal Society Open Science*, *2*(5), 150081+.

Panteli, M. & Mauch, M. (2015). Suitability of audio features for rhythmic and melodic description of world music styles. In *unpublished*, volume 00, (pp. 00–00).

Savage, P. & Brown, S. (2014). Mapping Music: Cluster Analysis Of Song-Type Frequencies Within And Between Cultures. *Ethnomusicology*, *58*, 133–155.

Savage, P. E. & Brown, S. (2013). Toward a new comparative musicology. *Analytical Approaches To World Music*, *2*, 148–197.

Weyde, T., Cottrell, S., Dykes, J., Benetos, E., Wolff, D., Tidhar, D., Gold, N., Abdallah, S., Plumbley, M. D., Dixon, S., Barthet, M., Mahey, M., Tovell, A., & Alencar-Brayner, A. (2014). Big data for musicology. *1st International Digital Libraries for Musicology workshop*, *00*, 00–00.

FLUID CONSTRUCTION GRAMMAR: A NEW COMPUTATIONAL PARADIGM FOR STUDYING MUSIC AND MEANING

Johan Loeckx
Artificial Intelligence Lab
Vrije Universiteit Brussel
jloeckx@ai.vub.ac.be

ABSTRACT

Fluid Construction Grammar is discussed as candidate computational representation for expressing musical knowledge, both syntactical as semantic. It is specifically intended to bridge the gap between computational efforts and musicology. It approaches the issue from a cognitive point of view and allows the expression of the relation between form and meaning. Also, it is embedded in the so-called "semiotic cycle", providing a powerful mechanism to create one and the same grammar both for parsing (analysis) and production (composition) of music.

1. INTRODUCTION

Music has drawn large interest from musicologists as well as computer scientists and Artificial Intelligence researchers in particular. Unfortunately, there is still a wide gap between these two branches of research: the analytical work done by theoretical musicologist and the experiments performed by computer scientists. For this reason, we present Fluid Construction Grammar, being its ultimate goal to bridge the distance between musicology and computer science, not unlike similar work done in the linguistic domain.

The key point here is not to suggest that music and language are (partly or completely) similar, but rather that the FCG paradigm makes a perfect candidate for describing and understanding music in a conceptual framework shared by musicologists and computer scientists. This paper is structured as follows. First, the concepts that make up the basis of FCG are introduced and its technical architecture sketched. Next, key detection is given as a prototype example to illustrate the power of the paradigm, wrapping up with conclusions.

2. FLUID CONSTRUCTION GRAMMAR

Above considerations exemplify the need of a computational framework to scrutinize musical phenomena. More in particular, the interaction between meaning and form in music, the social/cultural interactions that take place as well as the aspects of music acquisition, learning and evolution deserve more research (Jackendoff, 2009).

The approach taken in this work is similar to work done in the field of Computational Linguistics (Mitkov, 2005). Investigating the analogies between music and language is not new (Feld & Fox, 1994; Jackendoff, 2009) – nor is using principles of grammar to describe and represent musical structures (Lerdahl & Jackendoff, 1983). Unfortunately, most current grammar-based models of music are implicitly or explicitly founded on generative grammars that are at odds with the cognitive advances in music. Furthermore, the existing computational frameworks are primarily targeted at computer scientists, rather than being a useful supporting analysis tool to musicologists.

In this light, the author wants to introduce Fluid Construction Grammar (FCG) (Hoffmann & Trousdale, 2013), a computational framework that was conceived in 1998 as an underlying formalism for modelling language evolution using language games played by autonomous robots. Contrary to generative grammars, FCG is not intended for generating random sentences, but to provide a framework for modelling musical listening and composition processes. It has proven its power and flexibility repeatedly in the past in the domain of linguistics (Steels, 2011; Gerasymova et al., 2009; van Trijp, 2011). In a same mindset, the author believes the description and understanding of music should be embedded (computationally) in its environment to accommodate for social and cultural effects and interactions and to study the evolution of music using, for example, multi-agent simulation.

We will now summarize the distinctive features and principles that make FCG a promising candidate for studying music in a computational setting.

2.1 Expression of meaning

FCG is as transparent as possible with respect to the musical theory and meaning that a musicologist aims to explore, as long as the analysis adopts a *constructional perspective*. Construction grammars addresses grammar from a cognitive perspective and are as such based on the fundamental idea that language (and, in our case, music) models can be described by a collection of "form and meaning" pairs. Such a pair will be referred to as a *(coupled) feature structure* in the remainder of the text. One of the advantages of using constructions is the fact that musical knowledge is more or less "contained" within the scope of a constructions (although these constructions of course can interact).

It is important to point out that FCG does not make any assumptions about the nature of meaning in music or even about the (non)existence of a strict dichotomy between the syntactic and semantic characteristics of a musical construction. The "meaning of meaning" is indeed a controversial topic (Wiggins, 1998) and the specific semantic and syntactic characteristics are to be chosen freely by

the designer, according to a philosophical or musicological theory. For example, the semantic part of a construction can represent functions in functional harmony, cognitive affects like consonance and tension, or embedded in a embodied paradigm.

2.2 The semiotic cycle

Similar to language perception, it is assumed that a musical creator or listener of music goes through a series of processes, called the *semiotic cycle* (Steels & Baillie, 2003). When a musical utterance is perceived ("parsing"), the hearer must (1) first de-render the utterance by extracting observable features, (2) reconstruct the meaning using his own musical knowledge (grammar) and (3) lastly, interpret this meaning in terms of his own "world" or cognitive model. During the reverse process of "production", meaning is translated to an utterance by conceptualizing its cognitive model, and formulating the musical meaning to translate it to an utterance using its musical knowledge, to finally render a written or performed form. These tasks are typically not performed in a strictly sequential order, and often it is the case that there is more than one hypothesis being worked on, to express a particular meaning or interpret a musical utterance. Very interesting properties of FCG are the fact that it supports multiple hypotheses and ambiguity and is robust to not perfectly grammatical phrases. Also, *the same set of construction is employed in parsing as well as production*, allowing parsing an utterance into a semantic interpretation, and vice versa, using the same bi-directional rule set of constructions.

2.3 Technical details

One and the same core engine handles both parsing and production. During processing, a coupled feature structure called the *transient structure* is built that contains all the information needed for parsing or production, divided into a semantic and syntactic pole. Grammatical constructions or rules, are coupled feature structures as well and thus contain a semantic and syntactic pole as well. FCG uses as many as possible existing widely accepted notions in theoretical and computational linguistics, including (a) *feature structures* for the representation of musical symbolic information; (b) *abstract templates or constructions* for the representation of grammatical patterns; (c) *general operators* for building up syntactic and semantic structures, such as Merge in Chomsky's Minimalist Grammar Goldberg (1995) and Unify in Jackendoff's framework Jackendoff (2002).

2.4 Differences with other paradigms

For those familiar with existing work on generative grammars for music production and representation, it is crucial to understand the fundamental differences between FCG and more "traditional" context-free grammars. First, FCG marks no sharp distinction between idiomatic and general rules; also, there is a continuum in the hierarchy and domain of rules: constructions can involve different levels of abstraction. For example, the same rule can for example apply to one note or to a motif and there no distinction in the formulation or processing of melodic or harmonic entries. Third, FCG allows schematisation through variable binding and categorisation to make grammatical constructions more abstract. Finally, constructions can be combined – several constructions all matching with different parts of the musical phrase – or integrated, using hierarchical constructions that combine partial structures into larger wholes. Also, it is possible to overlay different constructions that each put additional constraints to the final phrase.

With respect to Constraint Satisfaction Programming techniques, the logic programming approach adopted in FCG should be seen as "parallel" to it rather than opposing. The advantages FCG with respect to CP's are the following. First, its *extensive flexibility*: the multiple viewpoints on all musical artefacts (notes, phrases,...) are not fixed or predefined beforehand but generated when constructions apply. In other words, constructions can create new viewpoints/aspects/musical representations during processing. They can also introduce (and eliminate) new variables for optimization (or search in our case). Similarly, variables can represent *hierarchical structures* of variable length & complexity. The organization in constructions furthermore allows us to *cope with the complexity* and finally, FCG allows parsing as well as production: constructions apply in a bi-directional way (if crafted accordingly).

3. APPLICATION: KEY DETECTION

A telling application of Fluid Construction Grammar is *key detection key and modulation*. Deriving the key of a motif/piece and pinpointing the exact point or region that modulation takes place, as well as the mechanisms behind it, is still cumbersome for computers. Methods like Harte et al. (2006) and Purwins et al. (2000), generally compute an abstract profile for each key based upon a compacted representation and try to match it with a stored prototype in a database to determine the key, for example using a nearest neighbour approach (Temperley, 2007). Though these "distributed" methods are successful, they provide no insight in *how modulation is performed* and show little musicality. One of the reasons is that they do not allow ambiguity to express the key, which is crucial when performing modulation.

Our proposal, therefore, starts from a cognitive perspective, based on earlier work by Longuet Longuet-Higgins & Steedman (1971) in which the expression of ambiguity is key. Its basic working is depicted in Fig. 1. For each voice, the scale to which its melody can belong to, is determined. Of course, there are often many possibilities. For this reason, every (vertical) chord is labelled by means of (harmonic) chord constructions in a second step. A sequence of measures is then said to belong to a particular key, if all notes in all voices belong to the scale connected to that key, and if the chord of the first degree (or, more strictly, a cadence like V-I) is part of this sequence.

An interesting point here is that this approach embraces ambiguity: a melody can potentially belong to different

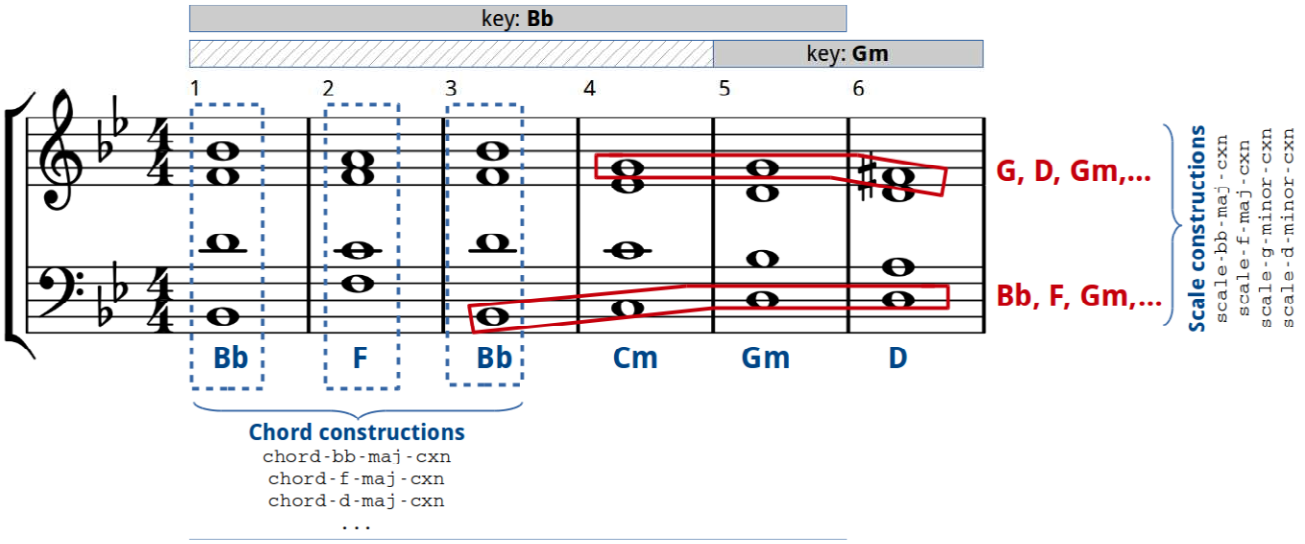


Figure 1: Chord and scale constructions applied to a 4-voice polyphonic melody. FCG allows ambiguity when determining the key, essential when describing modulation. Measure 5 can be in both the key of B-flat as G minor.

scales just like a part of a melody can belong to different keys. When more contextual information becomes available, the intentionally introduced ambiguity gradually dissolves. For example, once the key-construction concludes that the key is B-flat, the previous ambiguity in (melodic) scales can be eliminated using this information.

4. CONCLUSIONS

The formalism described in this paper allows the fine-grained expression of musical information through grammatical constructions that are either purely horizontal, vertical or a combination of both. The feature-based approach enables a multiple-viewpoint representation of music and the building of continuous hierarchies. These characteristics combined with the possibility to combine information across different levels of abstraction, provides a powerful framework for representing musical form as well as meaning. As such, the methodology suggested in this paper, attempts to bridge the gap between computer science and musicology by coming up with musically sensible formulations, rather than purely computationally constructed artefacts.

5. REFERENCES

Feld, S. & Fox, A. (1994). *Music and language*. Annual Review of Anthropology.

Gerasymova, K., Steels, L., & Van Trijp, R. (2009). Aspectual morphology of russian verbs in fluid construction grammar. In *Proceedings of the 31th Annual Conference of the Cognitive Science Society*, (pp. 1370–1375). Cognitive Science Society gerasymova-09a. pdf Google Scholar.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Harte, C., Sandler, M., & Gasser, M. (2006). Detecting harmonic change in musical audio. In *Proceedings of the 1st ACM workshop on Audio and music computing multimedia*, (pp. 21–26). ACM.

Hoffmann, T. & Trousdale, G. (2013). *The Oxford handbook of construction grammar*. Oxford University Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

Jackendoff, R. (2009). *Parallels and nonparallels between language and music*. Music Perception.

Lerdahl, F. & Jackendoff, R. (1983). *An overview of hierarchical structure in music*. Music Perception.

Longuet-Higgins, H. C. & Steedman, M. J. (1971). On interpreting Bach. *Machine intelligence*, *6*, 221–241.

Mitkov, R. (2005). *The Oxford handbook of computational linguistics*. Oxford University Press.

Purwins, H., Blankertz, B., & Obermayer, K. (2000). A new method for tracking modulations in tonal music in audio data format. In *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, volume 6, (pp. 270–275). IEEE.

Steels, L. (2011). *Design patterns in fluid construction grammar*, volume 11. John Benjamins Publishing.

Steels, L. & Baillie, J.-C. (2003). Shared grounding of event descriptions by autonomous robots. *Robotics and autonomous systems*, *43*(2), 163–173.

Temperley, D. (2007). *Music and probability*. The MIT Press.

van Trijp, R. (2011). Feature matrices and agreement: A case study for german case. *Design Patterns in Fluid Construction Grammar*, *11*, 205.

Wiggins, G. A. (1998). Music, syntax, and the meaning of "meaning". In *Proc. of First Symposium on Music and Computers*, (pp. 18–23).

A CULTURE-SPECIFIC ANALYSIS SOFTWARE FOR MAKAM MUSIC TRADITIONS

Bilge Miraç Atıcı
Bahçeşehir Üniversitesi
miracatici@gmail.com

Barış Bozkurt
Koç Üniversitesi
barisbozkurt0@gmail.com

Sertan Şentürk
Universitat Pompeu Fabra
sertan.senturk@upf.edu

ABSTRACT

Computational analysis of traditional music recordings often requires culture-specific problem definition and methodologies. Our previous efforts were directed towards developing technology for computational analysis of Turkish makam music which shares many common features with other maqam/makam traditions such as Arabic and Persian traditional music. This study presents an interactive tuning analysis tool developed with sufficient flexibility for parameter settings. We demonstrate that the tool can be effectively used for tuning analysis of Turkish music, Arabic music and Persian music once the appropriate settings are supplied by the user, using the interface of the tool.

1. INTRODUCTION

Each music tradition in the world has distinct melodic, rhythmic and timbral characteristics as well as semantic and cultural understandings (Tzanetakis et al., 2007). Any study done on a music culture needs to have a careful consideration of these musical concepts and how they are related to each other. Until recently, computational methods and tools were mainly aimed at studying euro-genetic musics (Serra, 2011). These approaches typically fail to address the research questions and the related computational tasks involving other music traditions (Bozkurt, 2008; Gulati et al., 2015). This brings a necessity to introduce culture-specific problem definitions and create novel methodologies to solve them.

On the other hand, computational research tools aimed at analyzing or modeling common musical concepts may be developed with flexibility to a certain extent. Such a feature is crucial in comparative studies between music traditions, where the methodologies should be capable of showing the similarities and/or differences between each music tradition consistently.

In this paper, we present a music analysis software that can be adjusted according to the musical properties of the studied music tradition(s). We focus on the so-called "tuning analysis" task, which is an important step in the melodic analysis of many music traditions. In this task the performed melodic intervals in one or more audio recordings are extracted and analyzed. The obtained model(s) can then be used for several related research problems such as tonic identification (Bozkurt, 2014), melodic mode recognition (Gedik & Bozkurt, 2010; Koduri et al., 2014), comparison of music theory and practice (Bozkurt et al., 2014) and expression analysis. The developed software mainly differs from similar tuning analysis software (such as TARSOS (Six & Cornelis, 2011)) by the flexibility of applying user specified culture-specific settings.

As a proof-of-concept we develop a toolbox named *MakamBox* and a separate settings creation tool. The settings tool is used to adapt the *MakamBox* to the culture-specificities of the studied music. We show that the toolbox and the setting tool can be used to analyze various music traditions such as Turkish-makam, Arabic-maqam and Persian music traditions.

The rest of the paper is organized as follows: in Section 2 previous problem definitions and related studies are briefly introduced, in Section 3 the methods useful for analyzing Turkish makam music are stated and in Section 4 the software implemented in this study is described in detail.

2. RELATED RESEARCH

There are considerable amount of MIR studies on euro-genetic music culture. However, the existing MIR methodologies are inadequate to analyze other music cultures due to certain characteristic differences (Serra, 2011). These characteristic differences have been discussed in recent studies. To illustrate: Indian art musics do not utilize descriptive scores, which can be used to analyze the musical traditions (Koduri et al., 2014). Arab-Andalusian music possess a special and unique musical richness as it has been influenced by several political and geographical developments throughout the history (Chaachoo, 2011). In Central African music there is no reference tuning, standardization of pitch frequencies and fixed pitch intervals (Moelants et al., 2006, 2007). Consequently, unlike the discrete pitch space representation in euro-genetic music, Central African music has continuous pitch space representation.

In this context, Turkish makam music holds a special position since it is a junction point for several different music traditions (Bozkurt et al., 2009) It has many similarities with other makam traditions, which are still alive in Middle East, some parts of Asia and Northern Africa, and contains many of aforementioned features of traditional music cultures (Bozkurt et al., 2014). Moreover, there are already comprehensive studies on makam histogram templates extraction (Bozkurt, 2008), tonic identification (Bozkurt, 2014), makam recognition (Gedik & Bozkurt, 2010), tuning analysis (Bozkurt et al., 2009), eval-

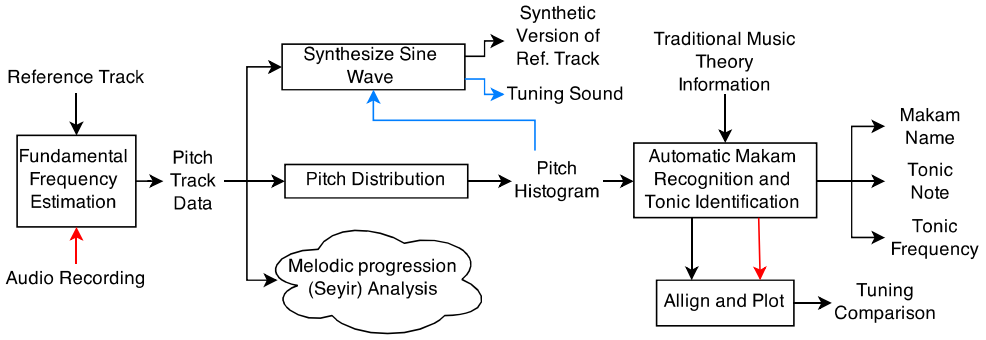


Figure 1: Block Diagram of *MakamBox*

uation of theory-performance mismatch and representation of pitch organization.

3. MELODIC ANALYSIS TASKS CONSIDERED FOR TURKISH MAKAM MUSIC

For the computational analysis of music traditions, several methods are needed to explore different aspects of the music. Mainly, analyzing the melodic and rhythmic features of a music culture provides a substantial amount of information. For this purpose, audio recordings, related metadata and machine readable scores can be used (Uyar et al., 2014). After the relevant data is obtained from the sources, various characteristics of the tradition can be studied. In this section, the methods used for analysis of Turkish makam music are briefly described.

Melody extraction is the first step of the analysis related to the melodic aspects. The result of the process is the pitch-time series data (pitch track). The pitch track provides important information about the audio recording such as the tuning (Bozkurt, 2012), the melodic progression (Şentürk et al., 2014) and embellishments (Özaslan et al., 2012). By using a pitch track, the pitch histogram can be obtained. Pitch histogram is computed as a distribution of the all pitch values in a recording.

A generic information about the scale features of a collection can be attained, by computing a pitch histogram from multiple audio files having a common characteristic. The pitch hierarchy and the melodic progression rules of the each makam result in certain distribution of the pitches. This approach has been used in many studies previously (Akkoç, 2002; Bozkurt, 2008; Karaosmanoğlu & Akkoç, 2003; Şentürk et al., 2013; Tzanetakis et al., 2003; Zeren, 2003). For instance, by combining the theoretical information and the pitch histograms of a group of recordings from a certain makam, the makam template histogram can be obtained. Makam template histograms of Turkish makam music theory can be computed by supervised learning as described in Bozkurt (2008). Automatic makam and tonic detection method provided by Bozkurt (2008) is based on matching the overall histogram of a certain audio recording with makam histogram template.

Pitch histograms have been used to analyze the tuning in makam music recordings in Bozkurt (2008); Bozkurt et al. (2009); Gedik & Bozkurt (2010). The comparison between two aligned histograms which are computed from two recordings in same makam is quite useful to understand tuning approach of different experts. In addition, pitch histograms can be used to compare analysis of student performance to reference track.

The tonic together with the *ahenk* [1] information provides information about the tuning used in a recording. This knowledge is important because there are several reference frequencies for tuning (e.g A $\approx$ 440 Hz for Mansur *ahenk*, D $\approx$ 440 Hz for Bolahenk *ahenk*) in practice of makam music tradition. If a user wants to practice by playing over a recording, a tuning difference between the user's instrument and the recording can be encountered. In such cases, key transposition of the recording is needed. The transposition can be performed by pitch shifting the recording with an amount of frequency difference between the estimated tonic and theoretical tonic frequency specified by the *ahenk* information and the makam.

Melody extraction methods may be inaccurate in some conditions. Some methods work well on monophonic recordings whereas they are not reliable on polyphonic recordings (Şentürk et al., 2014). In addition, some octave errors on the pitch track effects the accuracy of the pitch distribution. Synthesizing sine wave using a portion of pitch track and listening both original recording and synthetic version are quite useful to test accuracy of estimation. Furthermore, synthesizing sound using histogram peaks is helpful to tune the instrument as used in a reference track.

4. TOOLS

4.1 Analysis Software

To increase the usability of the existing computational studies of TMM, *Makam Aracı* (Bozkurt, 2014) has been developed for the musicians and the musicologists in the MATLAB [2] environment. *Makam Aracı* consists of some implementations of Turkish makam music studies. The tool-

[1] Diapasons are called *ahenk*s in Turkish Makam Music, which roughly specifies the length of ney that will serve as the reference mapping between frequencies and notes (similar to specifying that a score should be interpreted using a Bb instrument).

[2] http://www.mathworks.com/products/matlab/

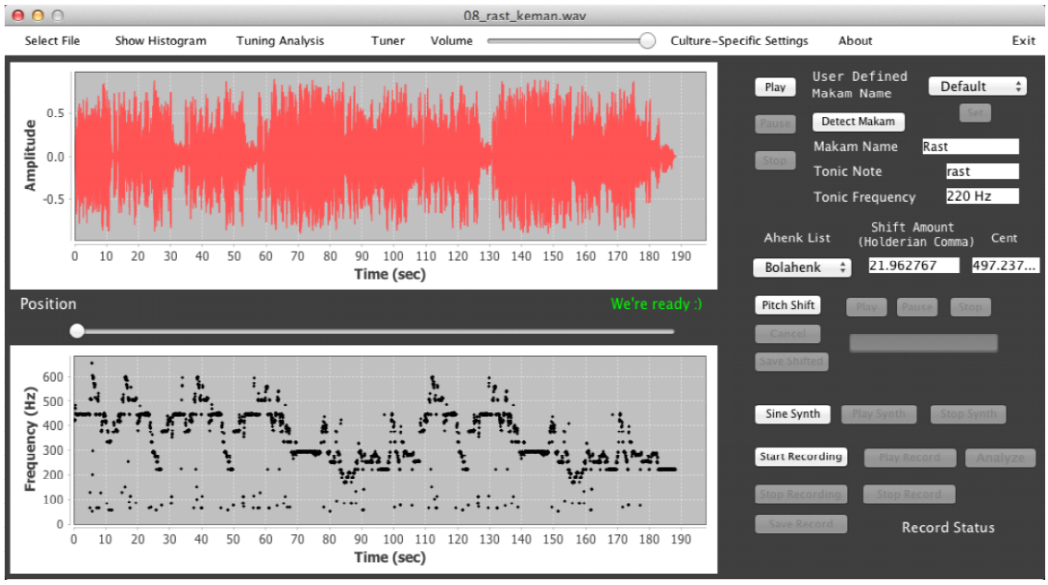


Figure 2: Screen capture of *MakamBox*, Culture-Specific Analysis Software

box has been used in several thesis studies in conservatories or musicology departments (Ekşi, 2011; Özek, 2011; Tan, 2011). Moreover, the methodologies implemented in this toolbox can be effective used with some adaptations for other makam music traditions (such as Arabic and Persian) by taking into consideration of the cultural similarities of Middle East, Asia and Northern Africa.

Although it is quite useful, *Makam Aracı* is not designed as a standalone application. It requires extra configuration steps (e.g. compilation of the YIN (De Cheveigné & Kawahara, 2002) algorithm, a need of MATLAB environment). In addition, *Makam Aracı* can be only used for the Turkish makam music culture due to hard-coded culture-specific music theory information.

To cope with these limitations and difficulties we have re-implemented *Makam Aracı*. We name the new software as *MakamBox*. *MakamBox* is developed in Java [3] to be able to deploy easily to researchers and musicians. The origin of our idea is to develop platform-independent software which does not require additional installation (excluding Java Virtual Machine, Java Runtime Environment or JRE).

Another advantage of *MakamBox* is the flexible analysis capabilities on different music cultures and traditions. As a consequence, all hard-coded information about TMM is removed from the software. Then, a new function that helps the user to load culture-specific settings file to software is added. The discussed culture-specific setting file which can be created via the newly developed tool, contains overall information (note and makam names, template histogram for each makam, *ahenk* information) about specified traditional music theory.

A snapshot of *MakamBox* is shown in Figure 2. The waveform and the pitch track of the reference recording are presented in the main window. The user can zoom in and out both of the graphics and loop-play the zoomed part. Moreover the estimated pitch track can be synthesized using a sine wave. The user can listen to the synthesized pitches corresponding to the peaks of the pitch histogram. The synthesized peaks can also be used a reference to tune musical instruments.

[3] http://www.java.com

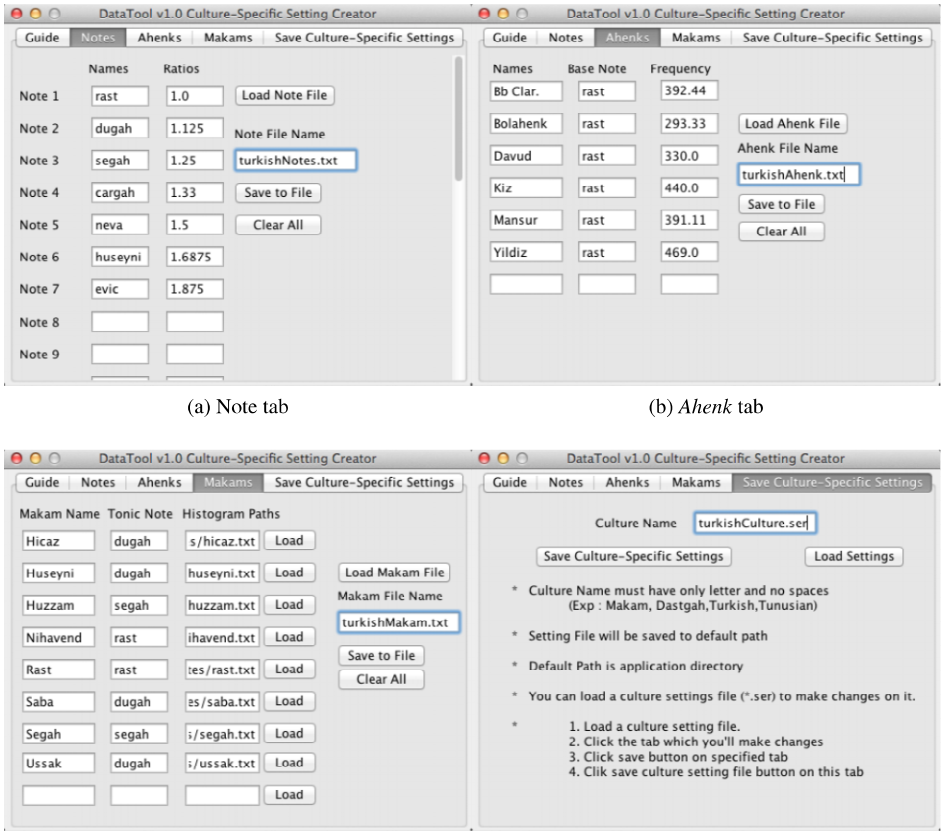
12TET Note Names	Frequency Ratios
C	1.0000
D	1.1225
E	1.2599
F	1.3348
G	1.4983
A	1.6818
B	1.8877
C (octave)	2.0000

Table 1: Note names and frequency ratios for 12TET

4.2 Culture-Specific Settings Creation Tool

DataTool is a software which is developed to create the settings file. *DataTool* needs 3 types of information. First one is about note information in traditional music theory. In this part, the user specifies the name of notes and frequency ratios with respect to the first note. This information serves as a theoretical reference, which helps initialization of intonation analysis. For example, frequency ratio of A5 with reference to A4 is 2. Frequency ratios of notes in 12 tone-equal-tempered (TET) and Turkish Makam Music (according to Arel theory (Arel, 1968)) are listed in Table 1 and Table 2 . On the notes tab in the software, *Save To File* button saves the notes names and frequency ratios to a text file which user can specify the name of the file. The snapshot of this part is shown in Figure 3a.

Second part is the *Ahenk Settings*. User will specify the *ahenk* name, reference note of each *ahenk* and frequency of the reference note. This data will be used for key trans-



(a) Note tab (b) *Ahenk* tab

(c) Makam tab (d) Culture-specific setting file tab

Figure 3: Screen captures of *DataTool* software

TMM Note Names (8 of 24)	Frequency Ratios
Rast	1.0000
Dügah	1.1250
Segah	1.2500
Çargah	1.3333
Neva	1.5000
Hüseyni	1.6875
Eviç	1.8750
Gerdaniye	2.0000

Table 2: Note names and frequency ratios for TMM (according to Arel)

position (when pitch shifting needs to be applied). *Save to File* button saves the *ahenk* names, reference notes and their frequencies to a text file which user can specify the name of the file. The snapshot of this part is shown in Figure 3b.

Third part is the *Makam Settings*. For each makam, a template histogram file is needed, which is computed/learned from recordings using another tool described in Bozkurt (2008). The user needs to specify the text file path containing the histogram. Also, each makam has a tonic note information, called *karar perdesi*. The shifting amount to transpose to user-specified *ahenk* with this note is calculated automatically. *Save to File* button saves the makam names, tonic notes and histogram file path to a text file which user can specify the name of the file. The snapshot of this part is shown in Figure 3c.

At the end of the process, the user needs to save all settings with given name to a Java SER [4] file. *Save Culture-Specific Settings* button saves the overall information about music culture to SER file, which user can specify the name. It will be used to set culture specific features of the *MakamBox*. There is also a *Load Settings* button. By clicking *Load* button the user can make some changes on settings file, which is created before. The snapshot of this part is shown in Figure 3d.

5. CONCLUSION

In this study, we presented an intonation analysis tool that can be adapted easily to various makam music traditions.

A new software, *DataTool*, is developed to expand the analysis capabilities of *MakamBox* on different music cultures and traditions. Interested readers are invited to author's web page [5] to watch our demonstration videos about creating culture-specific settings file and its usage.

[4] The file name extension which is used mostly for Java serializable object

[5] www.miracatici.com/makambox

6. ACKNOWLEDGMENTS

This work is supported by the European Research Council under the European Unions Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

7. REFERENCES

Akkoç, C. (2002). Non-deterministic scales used in traditional Turkish music. *Journal of New Music Research*, *31*(4), 285–293.

Arel, H. S. (1930, Reprint: 1968). *Türk Musikisi Nazariyatı*. Istanbul, Turkey: Hüsnütabiat Matbaası.

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

Bozkurt, B. (2012). A system for tuning instruments using recorded music instead of theory-based frequency presets. *Computer Music Journal*, *36*(3), 43–56.

Bozkurt, B. (2014). Pitch histogram based analysis of makam music in Turkey. In *Proceedings of Les Corpus de l'oralité*, Sampzon, France. Delatour France.

Bozkurt, B., Ayangil, R., & Holzapfel, A. (2014). Computational analysis of turkish makam music: Review of state-of-the-art and challenges. *Journal of New Music Research*, *43*(1), 3–23.

Bozkurt, B., Gedik, A., & Karaosmanoglu, M. (2009). Music information retrieval for Turkish music: problems, solutions and tools. In *Proceedings of IEEE 17th Signal Processing and Communications Applications Conference*, (pp. 804–807)., Malatya, Turkey.

Bozkurt, B., Yarman, O., Karaosmanoğlu, M. K., & Akkoç, C. (2009). Weighing diverse theoretical models on Turkish maqam music against pitch measurements: A comparison of peaks automatically derived from frequency histograms with proposed scale tones. *Journal of New Music Research*, *38*(1), 45–70.

Chaachoo, A. (2011). *La música Andalusí al-Ála: Historia, conceptos y teoría musical*. Córdoba, Spain: Almuzara.

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

Ekşi, O. (2011). Tarihi Kırkpınar güreş musikisi icracısının yetişme yetiştirme sürecinin değerlendirilmesi. Master's thesis, Trakya University, Edirne, Turkey.

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

Gulati, S., Serrà, J., & Serra, X. (2015). An evaluation of methodologies for melodic similarity in audio recordings of Indian art music. In *Proceedings of 40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (pp. 678–682)., Brisbane, Australia.

Karaosmanoğlu, M. & Akkoç, C. (2003). Türk musikisinde icra-teori birliğini sağlama yolunda bir girişim. In *10. Müz-Dak (Türk Müziği Dernek ve Vakıfları Dayanışma Konseyi) Sempozyumu*, Istanbul, Turkey. Istanbul Technical University.

Koduri, G. K., Ishwar, V., Serrà, J., & Serra, X. (2014). Intonation analysis of rāgas in Carnatic music. *Journal of New Music Research*, *43*(1), 72–93.

Moelants, D., Cornelis, O., Leman, M., Gansemans, J., De Caluwe, R., De Tré, G., Matthé, T., & Hallez, A. (2007). The problems and opportunities of content-based analysis and description of ethnic music. *International Journal of Intangible Heritage*, *2*, 57–68.

Moelants, D., Cornelis, O., Leman, M., Gansemans, J., De Caluwe, R. M., De Tré, G., Matthé, T., & Hallez, A. (2006). Problems and opportunities of applying data-& audio-mining techniques to ethnic music. In *Proceedings of 7th International Society for Music Information Retrieval Conference*, (pp. 334–336)., Victoria, Canada.

Özaslan, T. H., Serra, X., & Arcos, J. L. (2012). Characterization of embellishments in ney performances of makam music in Turkey. In *Proceedings of 13th International Society for Music Information Retrieval Conference*, (pp. 13–18)., Porto, Portugal.

Özek, E. (2011). *Türk müziğinde çeşni kavramı ve icra teori faklılıklarının bilgisayar ortamında incelenmesi*. Istanbul, Turkey: Proficiency in Arts Thesis, Haliç University.

Şentürk, S., Gulati, S., & Serra, X. (2013). Score informed tonic identification for makam music of Turkey. In *Proceedings of 14th International Society for Music Information Retrieval Conference*, (pp. 175–180)., Curitiba, Brazil.

Şentürk, S., Holzapfel, A., & Serra, X. (2014). Linking scores and audio recordings in makam music of Turkey. *Journal of New Music Research*, *43*(1), 34–52.

Serra, X. (2011). A multicultural approach in music information research. In *Proceedings of 12th International Society for Music Information Retrieval Conference*, (pp. 151–156)., Miami, Florida (USA).

Six, J. & Cornelis, O. (2011). Tarsos-a platform to explore pitch scales in non-western and western music. In *Proceedings of the 12th International Society for Music Information Retrieval Conference*, (pp. 169–174)., Miami, Florida (USA).

Tan, O. (2011). *Ney açkısının tarihi ve teknik gelişimi*. PhD thesis, Istanbul Technical University, Istanbul, Turkey.

Tzanetakis, G., Ermolinskyi, A., & Cook, P. (2003). Pitch histograms in audio and symbolic music information retrieval. *Journal of New Music Research*, *32*(2), 143–152.

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

Uyar, B., Atlı, H. S., Şentürk, S., Bozkurt, B., & Serra, X. (2014). A corpus for computational research of Turkish makam music. In *Proceedings of 1st International Digital Libraries for Musicology Workshop*, (pp. 57–63)., London, United Kingdom.

Zeren, M. A. (2003). *Müzik sorunlarımız üzerine araştırmalar*, volume 107. Istanbul, Turkey: Pan Yayıncılık.

DENSMORE REVISITED: CONTRAST DATA MINING OF NATIVE AMERICAN MUSIC

Kerstin Neubarth
Canterbury Christ Church University, Canterbury, United Kingdom
kerstin.neubarth@canterbury.ac.uk

1. INTRODUCTION

The American anthropologist Frances Densmore collected Native American music on field trips over several decades in the 20th century, and published collected songs with commentary and analysis as Bulletins of the Smithsonian Institution's Bureau of American Ethnology (Smithsonian Institution, 1971). While in many volumes "the songs of each tribe [were] compared with the total number of songs recorded and analyzed in other tribes" (Densmore, 1943, p. 181), other analyses list all groups individually.

These two approaches can be related to different comparison strategies in *contrast data mining*, a field of descriptive data mining which develops methods to discover differences between subsets or groups in data: *one-vs-all* strategies compare the examples of one group against examples of all other groups taken together (Novak et al., 2009); *separate grouping* strategies compare all groups simultaneously (Bay & Pazzani, 2001). Simeon & Hilderman (2008) expect different comparison strategies to lead to different findings. Existing studies in descriptive data mining of folk music generally employ a one-vs-all approach (Taminau et al., 2009; Conklin & Anagnostopoulou, 2011; Anagnostopoulou et al., 2013). The current study applies contrast data mining to Densmore's collection and assesses separate grouping and one-vs-all strategies for discovering differences between groups of folk songs.

2. CONTRAST DATA MINING

The task of applying contrast mining to Densmore's data consists of identifying musical features whose support differs significantly between groups in the corpus. Here a *feature* is an attribute-value pair; e.g. the value 'harmonic' of the attribute 'structure' describes songs in which contiguous accented tones form simple chord relations. Groups in the corpus are constituted by the songs of different Native American tribes. The (relative) support of a feature X in a group G is the percentage of folk songs in the group for which the feature is true: $P(X \mid G)$.

A feature is a *contrast feature* if it meets two requirements: significance and effect size (Bay & Pazzani, 2001). A candidate contrast feature is *significant* if support differences between groups are statistically valid, i.e. a statistical test such as a χ^2-test rejects the null hypothesis that the support is independent of group membership. A candidate contrast feature is *large* if the support deviation between groups meets a specified threshold θ. In this study effect size is measured by support ratio, which has been used in separate grouping (Simeon & Hilderman, 2008) and one-vs-all comparisons (Zhang et al., 2000; Conklin, 2010).

Separate grouping comparisons are based on pairwise evaluations of a feature's support between groups (Bay & Pazzani, 2001): a feature is a contrast feature if there is at least one pair of groups whose support ratio satisfies the effect size threshold

$$\max_{ij} \frac{P(X \mid G_i)}{P(X \mid G_j)} \geq \theta$$

and the difference in support across all groups is statistically significant according to the χ^2-test.

One-vs-all comparisons, on the other hand, select each group in turn as the target group and compare it against the rest of the corpus (Novak et al., 2009). For the purposes of this study a feature is considered a contrast feature based on one-vs-all comparison if there is at least one group for which the feature's support differs significantly from its aggregated support in all other groups:

$$\max_{i} \frac{P(X \mid G_i)}{P(X \mid \neg G_i)} \geq \theta$$

In this study the effect of the choice of comparison strategy on contrast mining results is assessed by determining the number of contrast features which are discovered by both separate grouping and one-vs-all comparisons, and the number of contrast features which are discovered by one of the two strategies but not the other.

3. DATA AND RESULTS

Densmore published quantitative analyses of folk music repertoires by eleven ethnic groups, covering melodic analysis (e.g. melodic ambitus of songs, tone material, initial melodic movement) and rhythmic analysis (e.g. initial metre, metrical position of the first note, metre changes). As the exact feature sets vary between different bulletins, a subset of groups and features was selected which can be mapped across bulletins. Some attribute values were merged, taking into account Densmore's own suggestions where available. The resulting data set comprises 940 folk songs organised in five groups. For this data set, occurrence counts for 13 attributes (65 features) were extracted from Densmore's analyses.

Table 1 shows the number of discovered contrast features at a significance level of 0.1 ($p < 0.0003$ with Bonferroni correction) for different support ratio thresholds.

		$\theta = 1.2$	$\theta = 2.0$	$\theta = 3.0$
Contrast features discovered by:	separate grouping	12	12	9
	one-vs-all	17	13	5
Contrast features discovered by:	both comparison strategies	11	9	3
	separate grouping but not one-vs-all	1	3	6
	one-vs-all but not separate grouping	6	4	2

Table 1: Number of significant ($p < 0.0003$) and large (at various support ratio thresholds θ) contrast features discovered in separate-grouping and one-vs-all comparisons.

For the χ^2-test to be considered valid, a minimum occurrence count of patterns is required in groups being compared (Bay & Pazzani, 2001): in the current study a feature should be true for at least four songs in a group.

Generally the number of contrast features decreases for higher thresholds; features which are no longer large at a support ratio threshold of 3-fold over-representation include ending on the third above the keynote or an ambitus of eleven or more tones. Larger effect size thresholds are favoured by separate grouping comparison, which evaluates the support between pairs of individual groups rather than the support in one group against the aggregated support over all other groups. For example, songs in which the tone material is based on a major triad are 6.6 times more frequent among songs of the Northern Ute than among songs of the Menominee ($p = 2.6e$-8), while they are 2.5 times more frequent among songs of the Northern Ute than across all other groups ($p = 4.1e$-6). On the other hand, one-vs-all comparison can alleviate the issue of small occurrence counts in single groups, which precludes evaluation of a candidate feature by the χ^2-test. For example, tunes transcribed with two or more accidentals are 6.2 times more frequent among the Menominee songs than among songs of the other groups ($p = 6.9e$-8); in the separate-grouping comparison this feature is not evaluated because in the Northern Ute sample (no songs with at least two accidentals) and the Mandan and Hidatsa sample (two songs with at least two accidentals) its occurrence count does not satisfy the condition for applying the χ^2-test.

4. CONCLUSIONS

Motivated by Frances Densmore's comparative analyses of North American native music, this study explores separate grouping and one-vs-all contrast mining strategies for discovering significant differences between groups of folk songs. Results confirm the suggestion of Simeon & Hilderman (2008) that different comparison strategies can lead to different mining results. Based on Densmore's analyses, the reported case study mined for single contrast features. Future work could analyse contrast sets of features, extracted from the recently completed digitised collection of Densmore's transcriptions (Shanahan & Shanahan, 2014), and could consider not only global features but also sequential patterns (Conklin, 2010).

5. REFERENCES

Anagnostopoulou, C., Giraud, M., & Poulakis, N. (2013). Melodic contour representations in the analysis of children's songs. In *Proceedings of the 3rd International Workshop on Folk Music Analysis (FMA 2013)* (p. 40-42). Amsterdam, Netherlands.

Bay, S. D., & Pazzani, M. J. (2001). Detecting group differences: mining contrast sets. *Data Mining and Knowledge Discovery*, *5*(3), 213-246.

Conklin, D. (2010). Discovery of distinctive patterns in music. *Intelligent Data Analysis*, *14*, 547-554.

Conklin, D., & Anagnostopoulou, C. (2011). Comparative pattern analysis of Cretan folk songs. *Journal of New Music Research*, *40*(2), 119-125.

Densmore, F. (1943). *Choctaw music*. Washington, DC: Smithsonian Institution Bureau of American Ethnology Bulletin 136.

Novak, P. K., Lavrač, N., & Webb, G. (2009). Supervised descriptive rule discovery: a unifying survey of contrast set, emerging pattern and subgroup mining. *Journal of Machine Learning Research*, *10*, 377-403.

Shanahan, D., & Shanahan, E. (2014). The Densmore collection of Native American songs: a new corpus for studies of effects of geography and social function in music. In *Proceedings of the 13th International Conference for Music Perception and Cognition (ICMPC 2014)* (p. 206-209). Seoul, South Korea.

Simeon, M., & Hilderman, R. J. (2008). Improving contrast set mining. In *Proceedings of the Doctoral Consortium* (Vol. 7, p. 1-4). University of Regina, Canada.

Smithsonian Institution. (1971). *List of publications of the Bureau of American Ethnology.* Washington, DC: Smithsonian Institution Bureau of American Ethnology Bulletin 200.

Taminau, J., Hillewaere, R., Meganck, S., Conklin, D., Nowé, A., & Manderick, B. (2009). Descriptive subgroup mining of folk music. In *2nd International Workshop on Machine Learning and Music at ECML/PKDD 2009 (MML 2009).* Bled, Slovenia. (w/o pages)

Zhang, X., Dong, G., & Kotagiri, R. (2000). Exploring constraints to efficiently mine emerging patterns from large high-dimensional data sets. In *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2000)* (p. 310-314). Boston, MA, USA.

APPROACHING VOCAL PRODUCTION IN WORLD'S MUSIC CULTURES – A MIXED METHODS STUDY BASED ON THE PHYSIOLOGY OF SINGING

Polina Proutskova
Goldsmiths, University of London
proutskova@googlemail.com

Christophe Rhodes
Goldsmiths, University of London
c.rhodes@gold.ac.uk

Tim Crawford
Goldsmiths, University of London
t.crawford@gold.ac.uk

Geraint Wiggins
Queen Mary University of London
geraint.wiggins@qmul.ac.uk

1. INTRODUCTION

This exploratory study of experts' tacit knowledge on vocal physiology was motivated by the ethno-MIR idea to formalise vocal production differences in different cultures. We wanted to teach computers to "listen" to singing: train computational models that would be capable of distinguishing between vocalisations based on differences in vocal production techniques. This interest originated from our involvement with Alan Lomax's Cantometrics experiment and its findings from 1960s. In this experiment, in which over 5000 recordings from more than 500 cultures were analysed, vocal production was expressed through parameters like vocal tension, volume, rasp, glottal shake, nasality, etc. Statistical analysis of these parameters together with anthropological descriptors of the societies where the recordings were made, produced some stunning outcomes. The most spectacular was the relationship between vocal tension (or vocal width) and the subordination of women: Lomax found that pre-marital sex sanction for women were consistently more severe in societies where tense, narrow singing was preferred. We took special interest in deconstructing and formalising the notion of "tense, narrow" vocal production as opposed to "open, wide and relaxed" singing in the hope to review Lomax's findings using a contemporary MIR approach.

We quickly realised that not only is it difficult to find commonly understood descriptors of singing in our everyday language. Singing experts have not succeeded in this task either. Even within a single culture that has been thoroughly studied, such as Western music, there is little agreement among professionals about basic terminology on vocal production (McGlashan, 2013; Garnier, 2007; Mitchell, 2003). Publications in English analysing vocal production in other cultures are rare (Födermayr, 1971; Bartmann 1994).

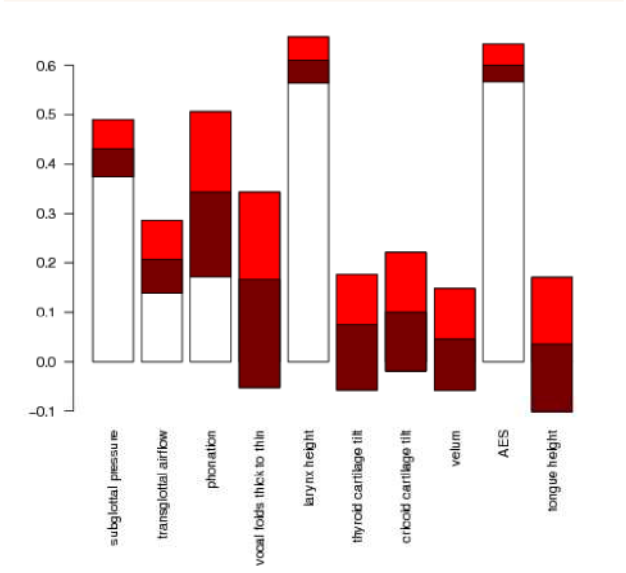
Singing teachers are concerned with vocal production routinely, but they often use idiosyncratic terminology based on their individual perception. Only recently has more objective language of physiology been introduced into practical teaching of singing, pioneered by the work of Jo Estill (Estill, 1979; Colton, 1981). Many progressive singing educators in the West have been inspired and influenced by Jo Estill's physiological approach (Sadolin, 2000; Soto-Morettini, 2006; Kayes, 2004). At the same time, her system is rarely borrowed in its entirety and her terminology has not become standard: Estill's physiological "building blocks" are a simplification of much more complex physiological processes and scientific evidence for some of them is inconclusive.

Medical professionals – phoniatricians, surgeons, otolaryngologists and speech language therapists - use rating systems such as GRBAS (Little, 2009) to assess vocal production, though these are primarily used to uncover voice disorders. Voice scientists have been studying singing for several decades (Sundberg, 1987; Howard, 1990; Titze, 2000). Though some important advances have been made, a comprehensive model of vocal production is yet to be developed.

physiological dimensions	range	scale	metrics
subglottal pressure	low to high	5-point	interval
transglottal airflow	low to high	5-point	interval
phonation breathy	present/absent	2-point	nominal
phonation pressed	present/absent	2-point	nominal
phonation neutral	present/absent	2-point	nominal
phonation flow	present/absent	2-point	nominal
vocal folds modal vs. falsetto	modal/falsetto	2-point	nominal
vocal folds vibration mode thick to thin	thick/mixed thicker/mixed/mixed thinner/thin	9-point	interval
larynx height	low to high	9-point	interval
thyroid cartilage tilt	vertical/slight tilt/tilted	5-point	interval
cricoid cartilage tilt	vertical/slight tilt/tilted	5-point	interval
velum	low to high	5-point	interval
aryepiglottic sphincter (size of vocal tract)	wide to narrow	5-point	interval
tongue height	low to high	5-point	interval
tongue compression	present/absent	2-point	nominal
position within chest register	low to high	5-point	interval
position within head register	low to high	5-point	interval

Table 1. Our preliminary ontology of vocal production.

The relationship between objective (physiological) and subjective (perceptual) descriptions of vocal production refers to one of MIR's biggest unresolved contradictions – the semantic gap (Wiggins, 2009). It follows the path of introducing a middle layer of objective, measurable descriptors in an attempt to model a high level characteristic (vocal production) via low-level information from the audio signal. Ability to extract vocal production from audio, if achieved, would benefit mainstream MIR (genre, mood, classification, etc) as well as MIR studies of non-Western musical cultures; it will also



constitute a significant contribution to other voice related disciplines.

Figure 1. Inter-participant agreement on the values of physiological descriptors for 13 participants. The picture shows confidence intervals for Krippendorff's alpha calculated via bootstrapping with 1000 iterations. Participants' confidence that was collected for each rating was also taken into account. The lower edge of the blue bar shows the 95% confidence level, the upper edge of the red bar – the 5% confidence level, the edge between the blue and the red

bars shows the actual value of alpha calculated from experts' ratings. If the confidence interval contains 0, then we can expect a higher than 5% probability that the given combination of ratings came about by chance and not through agreement. In contrast, where the confidence interval is well above zero, a tendency to agreement can be concluded, in this case for subglottal pressure, larynx height and AES.

2. THE STUDY

The aim of our study is to assess the viability of the physiological approach to modelling vocal production as well as to verify applicability and usefulness of our preliminary ontology of vocal production (Table 1). The study is based on interviews with vocal physiology experts and combines a qualitative and a quantitative approach (Bryman, 2006).
We chose eleven tracks from the Cantometrics dataset (see Chapter on vocal width in Lomax, 1977), all from different musical cultures. Nineteen physiologically stable fragments were extracted from the tracks, which were then used as entities of analysis in the interviews. We recruited 13 participants: otolaryngologists, speech language therapists, singing teachers. Participants' professional involvement with vocal physiology ranged from 10 to over 40 years. Three of them had a non-Western cultural background.
Interviews were structured and lasted from 90 minutes to several hours. Participants were asked to rate physiological dimensions from the preliminary ontology with which they were familiar; they were encouraged to explain their ratings, to point out complexities, to suggest better terms and approaches.

3. QUANTITATIVE ANALYSIS: RESULTS

Participants showed confidence in the majority of terms introduced in the preliminary ontology: only 20% of physiological dimensions were rated by less than 80% of participants. Most experts expressed general support for the ontology when asked open questions, only few additions were mentioned. Also, in the absolute majority of cases raters were satisfied with the choice of musical fragments to be analysed.
We chose Krippendorff's alpha as the measure of inter-participant agreement: it incorporates several other familiar estimators and can deal with partial data (Krippendorff, 2012).
Only for three physiological dimensions – subglottal pressure, larynx height and ari-epiglottic sphincter (AES) – did participants show a tendency to agreement (Figure 1). For all other dimensions agreement cannot be claimed.
This rather fuzzy picture demonstrates that experts of vocal physiology, who seem to be confident about the terminology that is used, in many cases do not agree about the values of the parameters they rate.

4. QUALITATIVE ANALYSIS: A PRELIMINARY PROBLEM CASES INVESTIGATION

In our talk we shall present a preliminary qualitative analysis of the interviews related to one particular musical example from Northeast Thailand. Lomax gives the following description of its context:

> "A country girl from this highly stratified, irrigation culture sings in a tense voice indicative of the sanctions and responsibilities that weigh upon S.E. Asian women. Her song thanks Buddha for the beauties of his creations—especially women. Mouth organ accompaniment." (Lomax 1977)

We shall analyse experts' perceptual descriptions of singing in the track and relate them to their physiology ratings. Bringing their views in their own words we shall look for insights to why their ratings differed in each instance. Their judgements of salience of physiological descriptors would give us another clue. Here is how differently it is described by three participants who agree that the singing sound in this example is rather relaxed (as opposed to being rated as tense in Cantometrics):

> "She's quite comfortable there, though it's very tight, a very small place, she's comfortable" (P11)
> "that's a very efficient sound" (P10)
> "If you think about the linguistic patterns and the voice quality that someone from that culture uses, probably for them it's not very tense. It sounds tense, to us, to a Western ear that sounds tense." (P15)

So far we have seen disagreement based on different interpretations of a term; genuine difficulties in determining a particular bit of a complex physiological process; changes in the vowel shape leading to physiological adjustments.
We expect to gain deeper understanding of discovered disagreement between experts on vocal physiology while the qualitative analysis of our study progresses.

REFERENCES

Bartmann, M. (1994). Rauhigkeiten in der Volksmusik in der Kanarischen Insel El Hierro. In *Böckner, M., editor, Berichte aus dem ITCM- Nationalkomitee Deutschland*, volume 3, Bamberg.

Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative research*, 6(1):97–113.

Colton, R. H. and Estill, J. A. (1981). Elements of Voice Quality: Perceptual, Acoustic, and Physiologic Aspects. In *Speech and language: advances in basic research and practice,* Lass, Norman J. (ed.), volume 5, pages 311–403. Academic Press, New York.

Estill, J. and Colton, R. (1979). The identification of some voice qualities. *The Journal of the Acoustical Society of America*, 65(S1).

Födermayr, F. (1971). Zu gesanglichen Stimmgenbung in der außereuropäischen Musik. Ein Beitrag zur Methodik der vergleichenden Musikwissenschaft. Stieglmayr, Wien.

Garnier, M., Henrich, N., Castellengo, M., Sotiropoulos, D., and Dubois, D. (2007). Characterisation of voice quality in western lyrical singing: from teachers' judgements to acoustic descriptions. *journal of interdisciplinary music studies*, 1(2):62–91.

Howard, D. M., Lindsey, G. A., and Allen, B. (1990). Toward the quantification of vocal efficiency. *Journal of Voice*, 4(3):205–212.

Kayes, G. (2004). Singing and the Actor. Routledge.

Krippendorff, K. (2012). Content analysis: An introduction to its methodology. Sage.

Little, M., et. al. (2009). Objective dysphonia quantification in vocal fold paralysis: comparing nonlinear with classical measures. *Journal of Voice*.

Lomax, A. (1968). Folk Song Style and Culture. Transaction Books, New Brunswick, New Jersey.

Lomax, A. (1977). Cantometrics: A Method of Musical Anthro- pology (audio-cassettes and handbook). Berkeley: University of California Media Extension Center.

McGlashan, J. (2013). What descriptors do singing teachers use to describe sound examples? Presented at *PEVOC 10 (Pan-European Voice Conference)* Prague, Czeck Republic.

Mitchell, H. F., T. Kenny, D., Ryan, M., and Davis, P. J. (2003). Defining 'open throat' through content analysis of experts' pedagogical practices. *Logopedics Phoniatrics Vocology*, 28(4):167–180.

Sadolin, C. (2000). Complete vocal technique. Shout Publishing Copenhagen, Denmark.

Soto-Morettini, D. (2006). Popular Singing: A Practical Guide To: Pop, Jazz, Blues, Rock, Country and Gospel. A&C Black.

Sundberg, J. (1987). The science of the singing voice. Illinois University Press.

Titze, I. R. (2000). Principles of voice production. National Center for Voice and Speech.

Wiggins, G. (2009). Semantic gap?? schemantic schmap!! methodological considerations in the scientific study of music. *Proceedings of IEEE AdMIRe*.

A NOVEL MUSIC SEGMENTATION INTERFACE AND THE JAZZ TUNE COLLECTION

Marcelo Rodríguez-López, Dimitrios Bountouridis, Anja Volk
Utrecht University, The Netherlands
{m.e.rodriguezlopez,d.bountouridis,a.volk}@uu.nl

ABSTRACT

In this paper we present MOSSA, an easy-to-use interface for mobile devices, developed to annotate the segment structure of music. Moreover, we present the *jazz tune collection* (JTC), a database of 125 Jazz melodies annotated using MOSSA developed specifically for benchmarking of computational models of melody segmentation. Each melody in the JTC has been annotated with segment boundaries by three human listeners, and segment boundary salience by two human listeners. We provide a light analysis of the inter-annotation-agreement of the annotations in the JTC, and also test the likelihood of the annotations been made using 'gap' related cues (large pitch intervals or inter-onset-intervals) and 'repetition' related cues (exact/approximate repetition of the beginning or ending of phrases).

1. INTRODUCTION

Music segmentation refers to a listening ability that allows human listeners to partition music into sections, phrases, and so on. Computational modelling of music segmentation is important for a number of fields related to Folk Music Analysis, such as Music Information Research (for tasks such as automatic music archiving, retrieval, and visualisation), Computational Musicology (for automatic or human-assisted music analysis), and Music Cognition (to test segmentation theories and more generally theories of musical structure).

Research in music segmentation modelling has been conducted by subdividing the segmentation problem into different tasks, most often segment boundary detection and segment labelling. Segment boundary detection is the task of automatically locating the time instants separating contiguous segments. Segment labelling is the task of categorising segments into equivalence classes. Generally, automatic segmentations are evaluated by comparing them to manual (human annotated) segmentations. In this paper we focus on the annotation of segment structure in melodies, which are of special interest in Folk Music Analysis.

1.1 Problem specification

Ideally, a melodic dataset used to test computational segmentation models should have the following two characteristics: first, it should comprise different styles and instrumental traditions, and second, each melody in the dataset should have been annotated by a relatively large number of human listeners.

However, at present most free and readily available annotated databases consist of vocal (mainly european) folk melodies. Furthermore, since the process of annotating segment structure in melodies is time consuming and laborious, participation to melody annotation initiatives is limited, and so melodic datasets are commonly annotated by a single expert annotator (or a small range of annotators that agree on a single segmentation).

Thus, there is a need for easy-to-use tools to avoid discouraging participation to melody annotation initiatives. Moreover, new melody databases are needed to account for stylistic and instrumental diversity when evaluating computational melody segmentation models.

1.2 Paper contributions

In this paper we present MOSSA (in §2) an interface for mobile devices which, aside of its portability, has a fast learning curve. Moreover, we present (in §3) and analyse (in §4) a database of 125 Jazz melodies annotated using MOSSA for benchmarking computational models of melody segmentation.

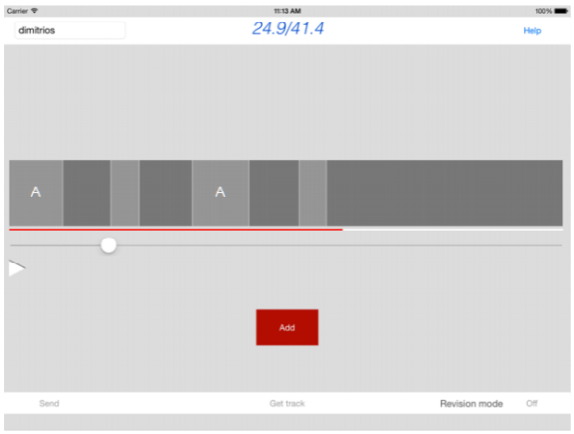


Figure 1: Screenshot of the MOSSA interface

2. MOSSA: MOBILE SEGMENT STRUCTURE ANNOTATION

Figure 1 shows a screenshot of the MOSSA interface. MOSSA is written in Objective-C for iOS. The code is available at http://www.projects.science.uu.nl/music/.

The main goals for the development of MOSSA, aside from portability, are (a) to avoid visual biases, and (b) to ensure a rapid learning curve. We elaborate into these two points below. (For a more detailed specification of the functionality of MOSSA the reader is referred to the documentation accompanying the code.)

2.1 Avoiding visual biases

Many segment structure annotation studies have used a score representation of the music to be annotated. This is specially true for melody segment annotation, e.g. (Thom et al., 2002; Pearce et al., 2010; Karaosmanoglu et al., 2014). Using a visual representation of musical content results in segment annotation biases. For instance, the geometry of score notation might influence the perception of boundary cues. For instance, score geometry might affect the perception of pitch interval distances and also make it easier to recognise repetitions of music material. This in turn might suggest the listener a particular segment structure that (s)he might not have been able to perceive without visual cues.

As seen in Figure 1 MOSSA avoids any visual representation of the music content, depicting music only as a time line. Different playback mechanisms are available for the user to easily examine whether the position of segment boundaries or its equivalent class labels are correctly annotated. For instance, if the user double taps a over a segment, playback starts from the leftmost boundary of the segment.

2.2 Ensuring fast learning

Most freely available interfaces for music annotation are rich in options, e.g. see (Li et al., 2006; Peeters et al., 2008; Cannam et al., 2006). However, the large number of options comes at the expense of user interaction simplicity, and hence may result in a relatively long and steep learning curve. MOSSA has been designed to minimise its learning time, by providing a clean and simple interface, and a visually intuitive way to annotate segment boundaries and label equivalent classes. For instance, as seen in Figure 1 boundaries can be inserted by simply pressing the 'add' button. Alternatively, boundaries can also be inserted by making a downwards swipe gesture over the block region representing the music.

The idea is that MOSSA is used by non-expert users, and then the annotations can be checked by experts in more advanced annotation interfaces, such as Sonic Annotator or Audacity.

3. THE JAZZ TUNE COLLECTION (JTC)

The JTC is a dataset of Jazz theme melodies constructed to evaluate computational models of melody segmentation. A list of global statistics describing the dataset is presented in Table 1.

Total number of melodies	125
Total number of notes	19419
Total time (in hours)	3.103
Approximate range of dataset (in years)	1880-1986
Total number of composers	81
Total number of styles	10

Table 1: Global statistics of the JTC

All melodies are available in MIDI. Each melody in the JTC is annotated with phrase boundaries (by three human listeners) and boundary salience (by two human listeners). [1] In Table 2 we present the total number of phrases and mean phrase lengths (with standard deviation values in parenthesis) per annotation.

Annotation	Number of Phrases	Mean Phrase Length	
		Notes	Seconds
1	1881	10.32 (4.85)	5.94 (4.85)
2	1701	11.42 (6.55)	6.57 (6.55)
3	1682	11.55 (5.78)	6.64 (5.78)

Table 2: Summary statistics of annotated phrases.

All segment boundaries and salience annotations were produced using MOSSA, and are provided in Audacity' label file format. The JTC also provides metadata for each melody. The metadata includes information of tune title, composer, Jazz sub-genre, and year of the tune's composition/release. The JTC dataset can be accessed at: `http://www.projects.science.uu.nl/music/`

3.1 JTC assembly

To assemble the JTC, we consulted online sources for Jazz album, tune, and composer rankings. [2] We employed a web-crawler to automatically collect MIDI and MusicXML files from a number of sources in the internet. (The majority were crawled from the now defunct *Wikifonia Foundation*. [3]) We cross referenced the rankings and the collected files, and selected 125 files trying to find a balance between tune ranking, composer ranking, sample coverage, and encoding quality. We describe the JTC's sample coverage (in terms of time periods and sub-genres) below, and discuss the encoding quality of the files in §3.3.

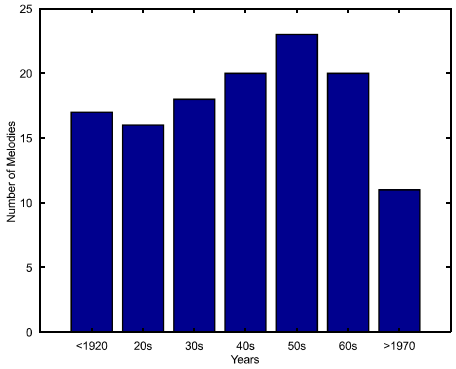


Figure 2: JTC: number of melodies per time period

The JTC can be divided in seven time periods (see Figure 2). Each time period contains between 11 and 23 tunes from representative sub-genres (see Figure 3) and influential composers or performers of the period. The year of

[1] We use the term 'boundary salience' to refer to a binary score that reflects the relative importance of a given boundary as estimated by a human annotator.

[2] The main sources consulted were: `www.allmusic.com`, `www.jazzstandards.com`, `en.wikipedia.org`

[3] `www.wikifonia.org`

release/composition, Jazz sub-genre, and composer metadata was obtained by consulting online sources. [4]

Class Label	Sub-Genre
C1	Bebop
C2	Big Band, Swing, Charleston
C3	Bossa Nova, Latin Jazz
C4	Cool Jazz, Modal Jazz
C5	Dixieland
C6	Early, Rag time, Folk Song
C7	Electric Jazz, Fusion, Modern
C8	Other
C9	Musical, Film, Broadway
C10	Post Bop, Hard Bop

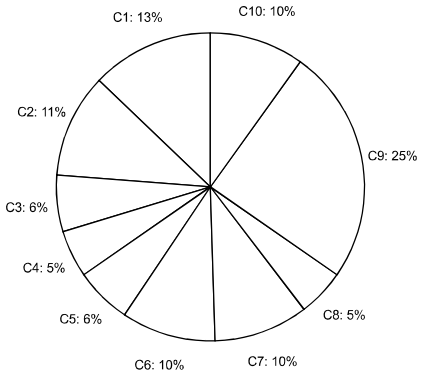


Figure 3: Distribution of sub-genres in the JTC

3.2 Melody encoding quality and corrections

From the 125 melodies making up the JTC, 64 correspond to perfomed MIDI files, 4 to manually encoded MIDI files, and 57 to manually encoded lead sheets in MusicXML format. In most cases the performed MIDI files encoded polyphonic music, so the melody was extracted automatically by locating the MIDI track labelled as 'melody'. [5]

All melodies were exported as MIDI files, using a resolution of 480 ticks-per-quarter-note. All melodies were inspected manually, and, if needed, corrected. Correction of the melodies consisted in adjusting note onsets, as well as removing ornamentation. Notated leadsheets from the Real Book series [6] were used as reference for the correction process. Is important to notice that not all ornamentation was removed, only that which was considered to severely compromise the understanding of segment structure. Also, while JTC melody encodings might contain information of meter, key, and dynamics, this information was not checked nor corrected, and thus its use as 'a priori' information for computational modelling of segmentation is discouraged.

3.3 Segment structure annotation process

For each melody, segment boundaries and salience were annotated by one amateur musician and one degree-level musician. These are referred to, respectively, as 'annotation 1 and 'annotation 2' in the Tables and Figures of this paper. For each melody there is also a third annotation of segment boundaries, produced by one of a group of extra annotators. This annotation is referred to as 'annotation 3' in the Tables and Figures of this paper.

The group of extra annotators consisted of 27 human listeners (18 male and 19 female), ranging from 20 to 50 years of age. In respect to the level of musical education of the extra annotators, 6 reported to at least play an instrument, 10 reported to having some degree of formal musical training, and 11 reported to having obtained a superior education degree in a music related subject. Moreover, extra annotators were asked to rate their degree of familiarity with Jazz (on a scale of 1 to 3, with 1 being the lowest, and 3 the highest), 12 annotators rated their familiarity as '1', 7 rated their familiarity as '2', and 8 rated their familiarity as '3'. Lastly, none of the extra annotators reported to suffering from any form of hearing impairment, and 2 reported having perfect pitch.

4. ANALYSIS OF PHRASE ANNOTATIONS

In this section we analyse the phrase annotations. In §4.1 we analyse two global properties of the annotated phrases: length and contours. In §4.2 we analyse inter-annotator-agreement using two different measures that score agreement. Finally, in §4.3 we check the vicinity of annotated phrases for evidence of two factors commonly assumed to be of high importance to segment boundary perception: *gaps* (in duration and pitch related information) and phrase start *repetitions* (also in duration and pitch related information).

4.1 Phrase Lengths and Contours

The mean phrase duration lengths presented in Table 2 and the box plots presented in Figure 4 show that the phrases of annotations 2 and 3 tend to be larger than those in annotation 1. Both boxes and whiskers of box plots 2 and 3 tend to be larger than those of box plot 1, indicating a larger spread skewed towards longer phrases. Furthermore, the notch of box plot 1 does not overlap with those of box plots 2 and 3, which indicates, with 95% confidence, that the difference between their medians is significant.

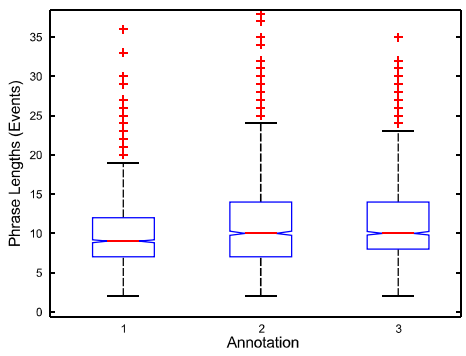


Figure 4: annotated phrase lengths

[4] in most cases `en.wikipedia.org` and `www.allmusic.com`

[5] If no such track was found the file was automatically filtered from the selection process.

[6] The Real Book editions used as reference for editing are published by `www.halleonard.com`.

To get further insights into these apparent preference for longer phrases, we consulted the degree-level musician of annotation 2 and some of the extra annotators for their choice of phrase lengths. The most common reply was that on occasion relatively long melodic passages suggested multiple segmentations, where phrases 'seemed to merge into each other' rather than having clear boundaries. For these passages the consulted annotators reported choosing to annotate just one long phrase with 'clear' boundaries rather than attempting to segment the melodic passage into multiple segments.

We also manually checked the outliers identified in Figure 4 for the presence of potential annotation errors. In most cases outliers simply correspond to melodic passages with high tempo and high note density, and are not particularly large in terms of time in seconds. Two examples of these type of outliers (common to of all annotations) are phrases in the melodies of Dexterity and Ornithology of Charlie Parker.

	Annotation		
Huron's Contour Classes	1	2	3
convex	33.86	35.10	36.15
descending	23.71	24.99	24.14
ascending	19.30	20.16	19.62
concave	19.99	16.34	17.06
ascending-horizontal	1.33	1.00	1.13
horizontal-descending	0.58	0.88	0.54
horizontal-ascending	0.37	0.59	0.48
descending-horizontal	0.48	0.47	0.42
horizontal	0.37	0.47	0.48

Table 3: Contour class classification of annotated phrases

We classified the annotated phrases in respect to their type of gross melodic contour using the contour types of Huron (1996). Table 3 shows the classification results, expressed as a percentage of the total number of phrases per annotation. The results show that all annotators agree in the ranking given to the four dominant contour classes, namely convex, descending, ascending, and concave (these four contour classes describe ∼96 percent of the phrases in each annotation). The ranking of the four dominant classes also matches the ranking obtained by Huron (1996), who performed phrase contour classification on ∼36000 vocal melodic phrases.

4.2 Inter-annotator-agreement (IAA) analysis

We checked the inter-annotator-agreement for each melody annotation using Cohen's κ (1960). Table 4 shows the mean pairwise agreement $\overline{\kappa}$, with standard deviation σ_κ in parenthesis. According to the scale proposed by Klaus (1980) the mean agreement on phrase boundary locations between annotations can be considered 'tentative', and according to the scale of Green (1997) it can be considered 'fair'. However, if for each melody we consider only the two highest κ scores, then $\overline{\kappa} = 0.86$, which can be considered by both the Klaus and Green scales as 'good/high'. Moreover, this 'best two' mean agreement also shows a substantial reduction in σ_κ. This indicates that, for any melody in the JTC, is likely that at least two segmentations have good agreement.

Annotation	$\overline{\kappa}$
1 vs 2	0.72 (0.22)
1 vs 3	0.71 (0.24)
2 vs 3	0.69 (0.26)
Best two	0.86 (0.15)

Table 4: Mean pairwise IAA ($kappa$)

Manual inspection of the boundary annotations showed that, even in cases when the annotators roughly agree on the total number of boundaries for a melody, constructing histograms of boundary markings results in clusters of closely located boundaries. We observed that these boundary clusters are in cases a side effect of dealing with ornamentation during segmentation (i.e. deciding whether grace notes, mordents, or fills should be part of one or another segment). We argue that boundary clusters are examples of 'soft' disagreement and should not be harshly penalised when estimating agreement.

The κ statistic does not take into account the possibility of, nor is able to provide partial scores for, points of 'soft' disagreement when estimating agreement. Hence, to investigate the effect of soft disagreement in the JTC we employed an alternative measure, namely the Boundary Edit Distance Similarity (B), recently proposed in (Fournier, 2013). One of the parameters of the B measure is a tolerance window (in notes). Within this tolerance window boundaries are given a partial score proportional to their relative distance. We tested the effect of soft disagreement by computing the B for each melody in the JTC using two tolerance levels: one note (giving score only to points strong agreement) and four notes (giving score also to points of soft agreement). We then computed whether the differences between the medians of the two sets of scores is statistically significant using a paired Wilcoxon Signed Rank test (WSRT). The results of this analysis are presented in Table 5. The WSRT confirms that the difference in medians is significant ($p < 0.001$), with medium effect size ($r = 0.41 - 0.47$). These results suggest that the number of points of 'soft' disarrangement is not negligible and it should be taken into consideration when benchmarking computational models of segmentation.

4.3 Analysis of Segment Boundaries

In this section we check annotated phrase boundaries and their immediate vicinity for the presence of two cues commonly assumed to be of high importance to segment boundary perception: melodic *gaps* and phrase start *repetitions*.

Melodic gaps can be defined as overly large changes in the temporal evolution of a given attribute used to describe a melody. Phrase start repetitions can be defined as an exact or approximate match of the attributes representing the starting point of two or more phrases. Our goal is to test to what extent gaps and repetitions can be considered a defining feature of the annotated phrase boundaries

Annotation	$\tilde{B}$ (tolerance = 1 note)	$\tilde{B}$ (tolerance = 4 notes)	WSRT
1	0.67	0.70	h: 1, Z: 4.54, $p < 0.001$, r: 0.41
2	0.62	0.67	h: 1, Z: 5.23, $p < 0.001$, r: 0.47
3	0.60	0.65	h: 1, Z: 5.23, $p < 0.001$, r: 0.47

Table 5: WSRT of B scores, tilde is used to denote the median, for the WSRT see Appendix A.1.

of the JTC. To that end, we make two complementary hypotheses: (a) the probability of detecting a gap at annotated phrase boundaries in a melody should be relatively high, which provides evidence that phrase boundaries often contain gaps, and (b) the probability of detecting a gap at non-boundary points in a melody should be relatively low, which provides evidence that gaps might be unique-to or distinctive-of phrase boundaries. The same pair of complementary hypotheses can be made for phrase start repetitions.

4.3.1 Computing per-melody detection probabilities

We compute the probability of detecting gaps/repetitions at/following boundaries

$$P_B = \frac{A_D}{A} \tag{1}$$

Where A_D is the number of annotated boundaries containing/preceding detected gaps/repetitions, and A is the total number of annotated boundaries in the melody. Likewise, we can compute the probability of detecting gaps/repetitions at/following non-boundaries

$$P_N = \frac{N_D}{N} \tag{2}$$

Where N_G is the number of non-boundaries containing/preceding detected gaps/repetitions, and N is the total number of non-boundaries in the melody.

4.3.2 Defining non-boundary points

We selected random non-boundary points with the following constraints: First, for each melody there should be an equal number of boundaries and non-boundaries. Second, non-boundary points should result in a set of segments of comparable length and standard deviation than that of the annotated phrases. With these two constraints, non-boundaries were drawn with uniform probability over eligible portions of the melody.

4.3.3 Gap analysis procedure

For gap detection we represent melodies as sequences of pitch or duration intervals. In this paper we measure pitch intervals (PI) in semitones, and measure duration using inter-onset-intervals (IOI) in seconds.

We classify (non-)boundaries as either containing or not containing a gap separately for PI and IOI using four different models of gap detection:

> T (Tenney & Polansky, 1980) in which a gap is detected if the interval at the (non-)boundary is larger than the intervals immediately preceding and following it.
>
> C (Cambouropoulos, 2001) in which a gap is detected if the interval at the (non-)boundary has a larger 'boundary strength score' than intervals immediately preceding and following it.
>
> R in which a gap is detected if the interval at the (non-)boundary is (a) equal or larger that four times the mode IOI of the melody, or (b) equal or larger than the mean PI of the melody plus one standard deviation.
>
> L in which a gap is detected if the interval at the (non-)boundary has (a) an IOI equal or larger than 1.5 seconds, or (b) a PI equal or larger than 9 semitones.

4.3.4 Repetition analysis procedure

For repetition detection we represent melodies as sequences of pitch intervals or inter-onset-interval ratios. We measure pitch intervals (PI) in semitones, and measure inter-onset-interval ratios (IOIR) in nats. [7]

We used the edit distance (Levenshtein, 1966) to compute similarity values S between the starting point of all phrases per melody. The similarity obtained per melody is normalised so that $S \in [0, 1]$. Pairwise phrase S values were computed separately for the PI and IOIR representation of the melody.

We define the start of a phrase according to the following rules. First, for an annotated segmented to be considered a valid phrase, we required segments to be longer than 2 intervals. Second, each valid phrase is divided in two (rounded to the nearest integer down) and the first half is used as a phrase start. If the first half is longer than 9 intervals truncation is applied. The maximum length of phrase start was chosen so that phrase starts are not longer than approximately the mean phrase size of the JTC (which according to Table 2 ranges between ∼10-11 notes).

For our experiments we classify phrase starts as either being repeated or not by considering three thresholds: similar ($S > 0.6$), closely similar ($S > 0.8$), and exact match ($S = 1$). [8]

4.3.5 Results

The results of the gap analysis are presented in Table 6. The results of the repetition analysis is presented are Table 7. To test if the differences between the medians of the obtained P_B and P_N scores are significant, we used once again the WSRT.

[7] The IOIR are computed using the formula and parameters proposed in (Wolkowicz, 2013, p. 45).

[8] For the exact match threshold we used the raw (not normalised) values of S.

Our results show that all annotations seem to roughly rank the tested cues in the same way. That is, IOI gaps are at the top of the ranking, with a P_B peaking at $\sim 0.95 - 1.00$, showing large and significant differences in respect to P_N scores. IOIR and PI repetitions are second, with P_B scores ranging between $\sim 0.30 - 0.66$, also showing relatively large and significant differences in respect to P_N scores. PI gaps are at the bottom of the raking, with P_B scores ranging $\sim 0.01 - 0.43$, showing in various cases non-significant differences in respect to P_N scores.

5. CONCLUSIONS

In this paper we have presented MOSSA a segment structure annotation interface for mobile devises. We have discussed some of the benefits of MOSSA in respect to existing segment structure annotation interfaces, such as its fast learning curve and avoidance of visual biases. In addition, we presented and analysed the *jazz tune collection* (JTC), a database of 125 Jazz melodies annotated using MOSSA developed specifically for benchmarking of computational models of melody segmentation. Our analysis of the JTC is aimed at investigating the inter-annotation-agreement of the annotations in the JTC, and also test the likelihood of the annotations been made using 'gap' related cues (large pitch intervals or inter-onset-intervals) and 'repetition' related cues (exact/approximate repetition of the beginning or ending of phrases).

6. REFERENCES

Cambouropoulos, E. (2001). The local boundary detection model (lbdm) and its application in the study of expressive timing. In *Proceedings of the international computer music conference*, (pp. 17–22).

Cannam, C., Landone, C., Sandler, M. B., & Bello, J. P. (2006). The sonic visualiser: A visualisation platform for semantic descriptors from musical signals. In *ISMIR*, (pp. 324–327).

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (1988). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Fournier, C. (2013). Evaluating text segmentation using boundary edit distance. In *Proc. of the 51st Annual Meeting of the Association for Computational Linguistics*, (pp. 1702–1712).

Green, A. M. (1997). Kappa statistics for multiple raters using categorical classifications. In *Proceedings of the 22nd annual SAS User Group International conference*, (pp. 1110–1115).

Huron, D. (1996). The melodic arch in western folksongs. *Computing in Musicology*, *10*, 3–23.

Karaosmanoglu, M. K., Bozkurt, B., Holzapfel, A., & Disiacik, N. D. (2014). A symbolic dataset of turkish makam music phrases. In *Proceedings of the 4th Folk Music Analysis Workshop (FMA)*, (pp. 10–14).

Klaus, K. (1980). Content analysis: An introduction to its methodology.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, (pp. 707–710).

Li, B., Burgoyne, J. A., & Fujinaga, I. (2006). Extending audacity for audio annotation. In *ISMIR*, (pp. 379–380).

Pearce, M., Müllensiefen, D., & Wiggins, G. (2010). Melodic grouping in music information retrieval: New methods and applications. *Advances in music information retrieval*, 364–388.

Peeters, G., Fenech, D., & Rodet, X. (2008). Mcipa: A music content information player and annotator for discovering music. In *ISMIR*, (pp. 243–248).

Tenney, J. & Polansky, L. (1980). Temporal gestalt perception in music. *Journal of Music Theory*, 205–241.

Thom, B., Spevak, C., & Höthker, K. (2002). Melodic segmentation: Evaluating the performance of algorithms and musical experts. In *Proceedings of the International Computer Music Conference (ICMC)*, (pp. 65–72).

Wolkowicz, J. M. (2013). Application of text-based methods of analysis to symbolic music.

A. APPENDICES

A.1 Wilcoxon Signed Rank test (WSRT)

Since the B scores can not be assumed to be normally distributed, we use the Wilcoxon Signed Rank test, which is a non-parametric alternative to the paired Students t-test, and gives the probability that two distributions of paired samples have the same median.

In this paper the results of the WSRT are reported using: h - test result (a value of 1 indicates the test rejects null hypothesis), Z - value of the z-statistic, p - p value, r - effect size. The effect size is computed as $r = {}^{Z}/\sqrt{N}$, where N is the total number of the samples. According to (Cohen et al., 1988), effect size values can be interpreted as small size if $r \leq 0.1$, medium size if $0.1 > r \leq 0.3$, large size if $0.3 > r \leq 0.5$, and very large size if $r > 0.5$.

IOI gaps				
Annotation	Gap Model	$\tilde{P}_A$	$\tilde{P}_N$	WSRT
1	T	0.95	0.20	h: 1, Z: 9.57, $p < 0.001$, r: 0.86
	C	0.94	0.21	h: 1, Z: 9.56, $p < 0.001$, r: 0.85
	R	0.67	0.04	h: 1, Z: 9.05, $p < 0.001$, r: 0.81
	A	0.56	0.02	h: 1, Z: 9.04, $p < 0.001$, r: 0.81
2	T	1.00	0.20	h: 1, Z: 9.64, $p < 0.001$, r: 0.86
	C	1.00	0.21	h: 1, Z: 9.63, $p < 0.001$, r: 0.86
	R	0.79	0.04	h: 1, Z: 9.14, $p < 0.001$, r: 0.82
	A	0.64	0.02	h: 1, Z: 9.08, $p < 0.001$, r: 0.81
3	T	0.96	0.20	h: 1, Z: 9.57, $p < 0.001$, r: 0.86
	C	0.96	0.20	h: 1, Z: 9.58, $p < 0.001$, r: 0.86
	R	0.78	0.04	h: 1, Z: 9.07, $p < 0.001$, r: 0.81
	A	0.61	0.04	h: 1, Z: 8.97, $p < 0.001$, r: 0.80
PI gaps				
Annotation	Gap Model	$\tilde{P}_A$	$\tilde{P}_N$	WSRT
1	T	0.25	0.27	h: 0
	C	0.42	0.36	h: 0
	R	0.29	0.11	h: 1, Z: 7.57, $p < 0.001$, r: 0.68
	L	0.01	0.01	h: 1, Z: 5.17, $p < 0.001$, r: 0.46
2	T	0.27	0.26	h: 0
	C	0.42	0.35	h: 0
	R	0.29	0.12	h: 1, Z: 6.78, $p < 0.001$, r: 0.61
	L	0.01	0.01	h: 1, Z: 5.33, $p < 0.001$, r: 0.48
3	T	0.29	0.26	h: 0
	C	0.43	0.34	h: 1, Z: 2.80, $p < 0.01$, r: 0.25
	R	0.27	0.10	h: 1, Z: 6.72, $p < 0.001$, r: 0.60
	L	0.01	0.01	h: 1, Z: 4.91, $p < 0.001$, r: 0.44

Table 6: Gaps at annotated boundaries and random boundaries, tilde is used to denote the median, for the WSRT see Appendix A.1.

Repetition of Phrase Beginning: IOI Ratio (IOIR)				
Annotation	Threshold	$\tilde{P}_A$	$\tilde{P}_N$	WSRT
1	$S > 0.6$	0.66	0.42	h: 1, Z: 8.56, $p < 0.001$, r: 0.77
	$S > 0.8$	0.50	0.25	h: 1, Z: 8.57, $p < 0.001$, r: 0.77
	$S = 1$	0.33	0.18	h: 1, Z: 7.65, $p < 0.001$, r: 0.68
2	$S > 0.6$	0.63	0.41	h: 1, Z: 8.56, $p < 0.001$, r: 0.77
	$S > 0.8$	0.50	0.26	h: 1, Z: 8.71, $p < 0.001$, r: 0.78
	$S = 1$	0.38	0.18	h: 1, Z: 7.49, $p < 0.001$, r: 0.67
3	$S > 0.6$	0.64	0.40	h: 1, Z: 7.92, $p < 0.001$, r: 0.71
	$S > 0.8$	0.50	0.27	h: 1, Z: 7.60, $p < 0.001$, r: 0.68
	$S = 1$	0.30	0.20	h: 1, Z: 6.90, $p < 0.001$, r: 0.62
Repetition of Phrase Beginning: Pitch Interval (PI)				
Annotation	Threshold	$\tilde{P}_A$	$\tilde{P}_N$	WSRT
1	$S > 0.6$	0.59	0.38	h: 1, Z: 8.46, $p < 0.001$, r: 0.76
	$S > 0.8$	0.46	0.23	h: 1, Z: 8.79, $p < 0.001$, r: 0.79
	$S = 1$	0.33	0.17	h: 1, Z: 7.76, $p < 0.001$, r: 0.69
2	$S > 0.6$	0.60	0.35	h: 1, Z: 8.55, $p < 0.001$, r: 0.76
	$S > 0.8$	0.50	0.24	h: 1, Z: 8.75, $p < 0.001$, r: 0.78
	$S = 1$	0.33	0.18	h: 1, Z: 7.37, $p < 0.001$, r: 0.66
3	$S > 0.6$	0.57	0.38	h: 1, Z: 7.74, $p < 0.001$, r: 0.69
	$S > 0.8$	0.43	0.25	h: 1, Z: 7.93, $p < 0.001$, r: 0.71
	$S = 1$	0.29	0.20	h: 1, Z: 6.84, $p < 0.001$, r: 0.61

Table 7: Repetitions at annotated and random phrase beginnings, tilde is used to denote the median, for the WSRT see Appendix A.1.

CLASSIFYING INSTRUMENTS WITH TIMBRAL FEATURES : APPLICATION TO ETHNOMUSICOLOGICAL RECORDINGS

Jean-Luc Rouas
LaBRI - CNRS, Université de Bordeaux,
33405 Talence, France
jean-luc.rouas@labri.fr

Dominique Fourer
IREENA - Université de Nantes,
44602 Saint-Nazaire, France
dominique.fourer@univ-nantes.fr

1. INTRODUCTION

The term "timbre" encompasses a set of auditory attributes of sound events in addition to pitch, loudness, duration, and spatial position. Automatic timbre characterization of audio signals can help to measure similarities between sounds and is of interest for automatic or semi-automatic databases indexing. In this presentation we evaluate an instrument classification method based on timbre features on the well studied Iowa musical instruments database, showing results comparable with those of state-of-the-art methods. Our method is then evaluated on the DIADEMS database, composed of worldwide nonwestern instruments audio recordings. We finally discuss on the practicality of this solution for automatic indexing of an ethnomusicological database.

2. DATABASES

2.1 IOWA Instruments database

The Iowa musical instruments database Fritts (1997) is a freely available database. The database includes strings, more winds and brass, and a Steinway piano - commom western instruments. Since 1997, these recordings have been freely available on the web [1] and may be downloaded and used for any projects, without restrictions. These have been used in over 270 published research articles and books. Some statistics on this database are given in Table 1.

Instrument class	Duration (s)	#
reed/flute and brass	5951	668
struck strings	5564	646
plucked strings	5229	583
bowed strings	7853	838
Total	24597	2735

Table 1: Description of the Iowa instruments database with duration and number of 10 s. segmented excerpts

2.2 DIADEMS database

The DIADEMS database [2] is composed of ethnomusicological field recordings involving various acoustic conditions (i.e. no recording studio) and from places all around the world. It contains more than 7000 hours of audio data from 1932 to nowadays on different supports from vinyl discs to digital recordings. Most of the musical instruments which compose this database are non-western and can be uncommon while covering a large range of musical instrument families (see Figure 1(a)). Among uncommon instruments, one can find the Ngbaka harp, oscillating bamboos, struck machete and struck girder. In this paper, we restricted our study to the solo excerpts (where only one monophonic or polyphonic instrument is active) to reduce the interference problems which may occur during audio analysis. A description of this selected CREM sub-database is presented in Table 2.

Class name	Duration (s)		#	
aerophones-blowed	1383		146	
cordophones-struck	357		37	
cordophones-plucked	715	1229	75	128
cordophones-bowed	157		16	
idiophones-struck	522		58	
idiophones-plucked	137	753	14	82
idiophones-clinked	94		10	
membranophones-struck	170		19	
Total	3535		375	

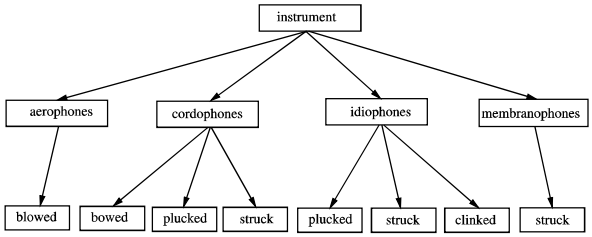
Table 2: Content of the CREM sub-database with duration and number of 10 s. segmented excerpts

3. TAXONOMY

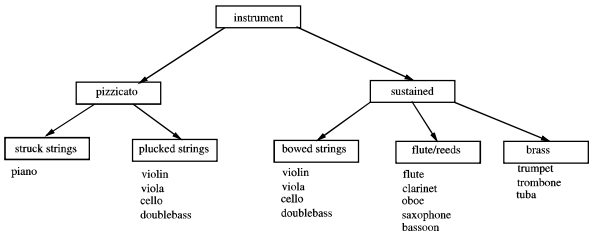
In this study, we use two databases which can be annotated using different taxonomies. Due to its diversity, the CREM database was only annotated using the Hornbostel and Sachs taxonomy v. Hornbostel & Sachs (1961) (T1) illustrated in Figure 1(a) which is widely used in ethnomusicology. This hierarchical taxonomy is general enough to classify uncommon instruments and conveys information about sound production materials and playing styles. From an another hand, the Iowa musical instruments database Fritts (1997) used in our experiments was initially annotated using a musician's instrument taxonomy (T2) as proposed in Peeters & Rodet (2002) and illustrated in Figure 1(b). The Iowa database is only composed of aerophones and cordophones instruments. If we consider the playing style, only 4 classes are represented if we apply T1 taxonomy to the Iowa database.

[1] http://theremin.music.uiowa.edu/MIS.html

[2] CREM audio archives freely available online at: http://archives.crem-cnrs.fr/



(a) Hornbostel and Sachs taxonomy (T1)



(b) Musician's instrument taxonomy (T2)

Figure 1: Taxonomies used for the automatic classification of musical instruments as proposed by v. Hornbostel & Sachs (1961) (a) and Peeters & Rodet (2002) (b).

Correspondence between classes of the two taxonomies can be assessed as : aerophone-blowed $\leftrightarrow$ reed/flute and brass, cordophone-strucked $\leftrightarrow$ struck strings, cordophone-plucked $\leftrightarrow$ plucked strings, cordophone-bowed $\leftrightarrow$ bowed strings.

4. CLASSIFICATION FRAMEWORK

In this study, we focus on the sound descriptors from the timbre toolbox [3] and detailed in Peeters et al. (2011). There is a total of 164 descriptors. To normalize the duration of analyzed sound, we separated each excerpt in 10-seconds length segments without distinction of silence or pitch events.

Inertia Ratio Maximisation using features space projection (IRMFSP) was first proposed in Peeters (2003) to reduce the number of descriptors used by timbre classification methods. It consists in maximizing the relevance of the descriptors subset for the classification task while minimizing the redundancy between the selected ones.

The goal of Linear Discriminant Analysis (LDA) Anderson (1958) is to find the best projection or linear combination of all descriptors which maximizes the average distance between classes (inter-class distance) while minimizing distance between individuals from the same class (intra-class distance). Each instrument class is modeled into the projected classification space resulting from the application of LDA.

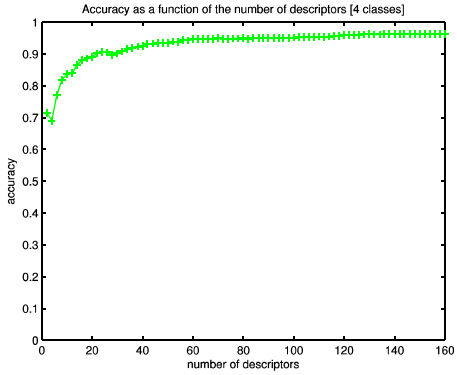


Figure 2: Iowa database classification (4 classes) - classification rate vs. number of features

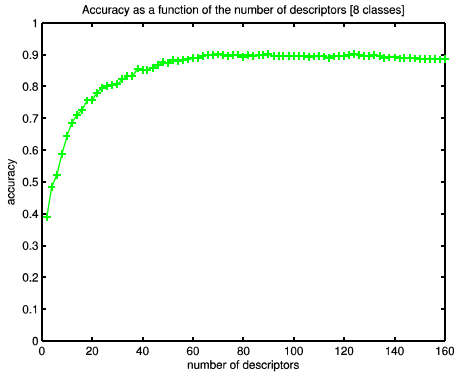


Figure 3: DIADEMS database classification (8 classes) - classification rate vs. number of features

5. EXPERIMENTS AND RESULTS

5.1 Iowa database

5.2 DIADEMS database

6. CONCLUSION AND FUTURE WORK

We applied a computationally efficient automatic timbre classification method which was successfully evaluated on an introduced diversified database using an ethnomusicological taxonomy. This method obtains good classification results ($> 80\%$ of accuracy) for both evaluated databases which are comparable to those of the literature. However, the cross-database evaluation shows that each database cannot be used to infer a classification to the other. This can be explained by significant differences between these databases. Interestingly, results on the merged database obtain an acceptable accuracy of about 70%.

7. ACKNOWLEDGEMENTS

This research was partly supported by the French ANR *(Agence Nationale de la Recherche)* DIADEMS *(Description, Indexation, Accès aux Documents Ethnomusicologiques et Sonores)* project (ANR-12-CORD-0022).

[3] MATLAB code available at `http://www.cirmmt.org/research/tools`

8. REFERENCES

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. New York, USA: Wiley-Blackwell.

Fritts, L. (1997). Musical instrument samples. Univ. Iowa Electronic Music Studios. [Online]. Available: http://theremin.music.uiowa.edu/MIS.html.

Peeters, G. (2003). Automatic classification of large musical instrument databases using hierarchical classifiers with intertia ratio maximization. In *115th convention of AES*, New York, USA.

Peeters, G., Giordano, B., Susini, P., Misdariis, N., & McAdams, S. (2011). The timbre toolbox: Extracting audio descriptors from musical signals. *Journal of the Acoustical Society of America*, *130*(5), 2902–2916.

Peeters, G. & Rodet, X. (2002). Automatically selecting signal descriptors for sound classification. In *Proc. ICMC*, Göteborg, Sweden.

v. Hornbostel, E. & Sachs, C. (1961). The classification of musical instruments. *Galpin Society Journal*, *3 (25)*, 3–29.

BODY VIBRATIONS AS SOURCE OF INFORMATION FOR THE ANALYSIS OF POLYPHONIC VOCAL MUSIC

Frank Scherbaum
UP Transfer, University of Potsdam
fs@geo.uni-potsdam.de

Wolfgang Loos
Berlin University of the Arts
wolfgang.loos@udk-berlin.de

Frank Kane
Kane.frank@gmail.com

Daniel Vollmer
Institute of Earth- and Environmental Sciences, University of Potsdam
daniel@geo.uni-potsdam.de

1. INTRODUCTION

Recordings have always played a central role in ethnomusicology. Although recording techniques and recording strategies have changed considerably over the course of time, often triggered by technological advances (e. g. Arom, 1976), the focus of interest in the past has been predominantly on acoustical sound and more recently on video. In the present study we are investigating the use of body vibrations of singers as an additional source of information for the analysis of polyphonic vocal music. This was motivated by the observation that body vibrations generated during the phonation process contain an enormous amount of information of analytical interest (e. g. regarding the vocal sound spectrum) while at the same time being essentially unaffected by the signals of other singers. As a consequence, individual voices but also the interaction of voices can be analyzed in great detail using digital signal processing tools. As a demonstration of this concept, we are presenting examples of the application of pitch analysis software for the documentation of the individual voices in different types of polyphonic oral music.

2. MEASUREMENT OF BODY VIBRATIONS

The investigation of phonatory vibrations in singers is not new (e. g. Kirikae et al., 1964), but for most of the second half of the last century the main interest in this topic seems to have been related to its use as a feedback signal for phonatory control (Sundberg, 1979, 1983, 1992; Coleman, 1988; Verrillo, 1992). The results in this context fell short of the expectations, however, (McAngusTodd, 1993) which may explain the decrease in the number of recent publications (naturally with exceptions, e. g. Takada, 2014) on this subject. In a medical context, however, e. g. related to the health effects of body vibrations (Wikström et al., 1994), for understanding the role of bone conduction in the hearing process (e. g. von Békésy, 1960, Henry & Letowski, 2007), or in the context of speech pathology, e. g. the measurement of the sound pressure level (SPL) of the voice, the investigation of body vibrations have been of considerable interest (Milutinović et al., 1997, Svec et al., 2005; Lamarche et al. 2008).

Body vibrations have also received attention in the context of communication in noisy environments (Rash et al., 2009, Heracleous et al., 2004) and most recently in the context of activity recognition and life-logging (Rahman et al. 2014, Yatani & Trong, 2014). The present work started from yet a different perspective, namely as part of a wider study aimed at the application of seismological analysis techniques to the study of the generation and propagation of audible acoustical signals in the human body. In this context we started a still ongoing comparison of the performance of a variety of piezo-based accelerometers, pick-ups, as well as bone and larynx microphones for recording body vibrations at different body locations. Fig. 1 shows the results of recordings of one of us (F.K) singing the Abkhazian song *Varado* recorded on slightly modified Albrecht AE 38 S2 larynx microphones (at larynx and ankle) and a NEC VS-BV 201 sensor taped to the toenail.

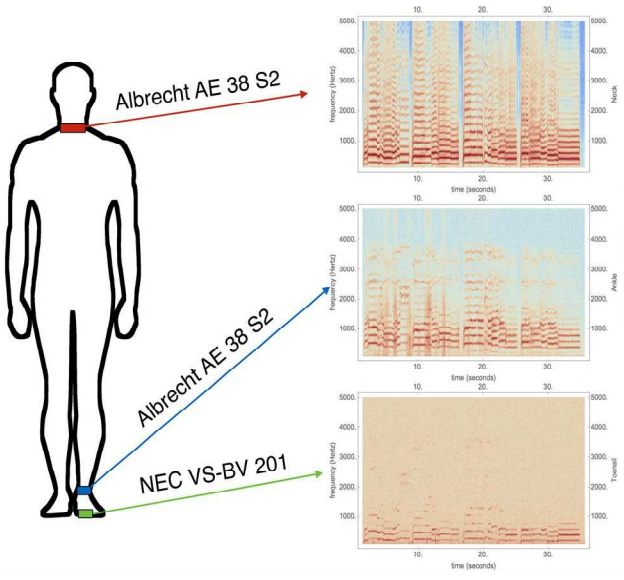


Figure 1. Sonograms of neck, ankle and toenail recordings (from top to bottom) of the beginning of the song *Varado*.

Acoustically, the song is clearly recognizable on playback from all three recordings, although the signal to-noise ratio decreased strongly for the toenail recording. The latter is partially attributed to poor coupling of the sensor.

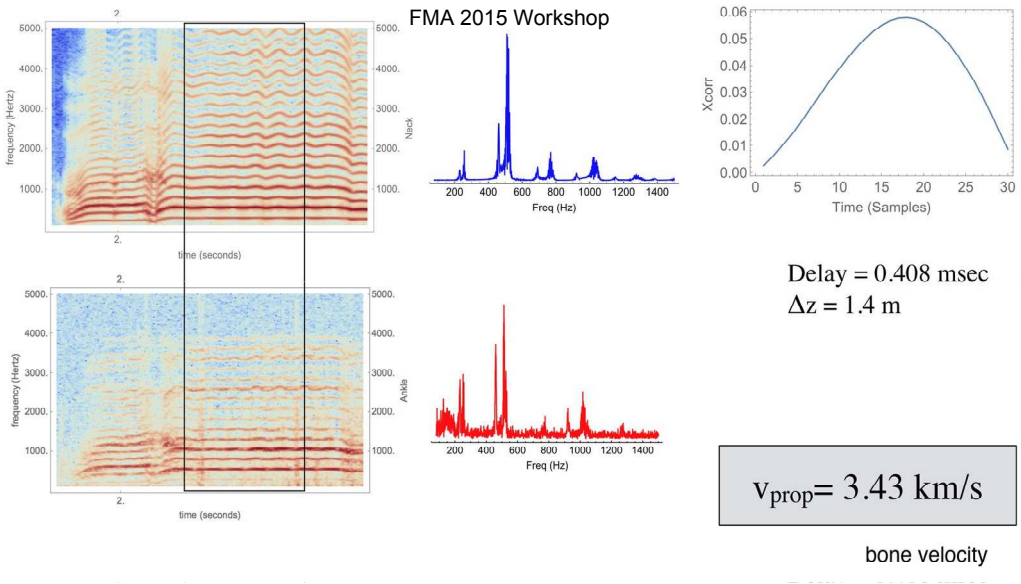


Figure 2. Sonograms (left panel), spectra (central panel) and cross-correlation function (right panel) between neck and ankle recordings of the beginning of the song *Varado*.

The neck and the ankle signals show a pronounced phase delay which was estimated by cross-correlation to be roughly 0.41 msec. This, based on the distance between the two microphones of 1.4 m, corresponds to a propagation velocity between ankle and neck of roughly 3.4 km/s (Fig. 2). Obviously, the ankle signal has propagated primarily through the bones. The relative depletion of high frequency components in the ankle signal (Figs. 1 and 2) might also be useful as a measurement of bone consistency (attenuation).

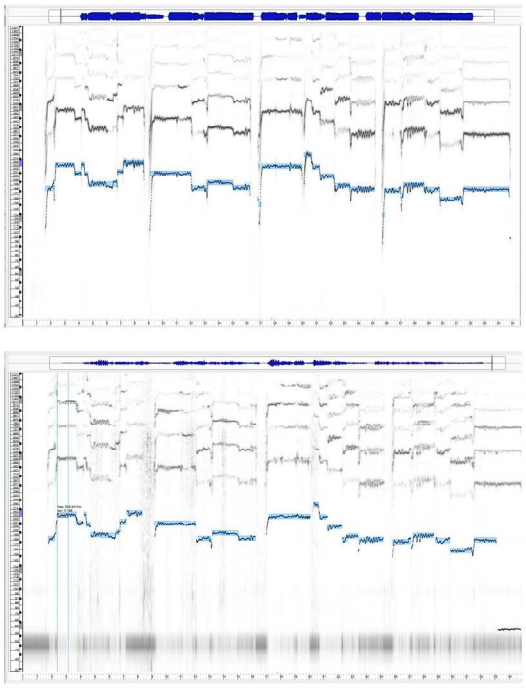


Figure 3. Automatic pitch recognition of the neck (top panel) and the ankle recording (bottom panel) of the song *Varado*, using the PYIN algorithm as implemented in the Tony software (Mauch and Dixon, 2014; Mauch et al., 2014, 2015).

Encouraged by the acoustic quality of the larynx microphone recordings at both the neck and the ankle position, we investigated the performance of automatic pitch analysis on these recordings (Fig. 3). For this purpose we used firstly the free Tony software (Mauch et al., 2014, 2015) and secondly the commercial Melodyne3 software package (Celemony Software GmbH). Melodyne3 contains a monophonic as well as a polyphonic pitch tracking algorithm without being able to output the results however. This limits the applicability of this software in the present context to illustration purposes. The Tony software on the other hand contains the well-documented PYIN algorithm (Mauch and Dixon, 2014) and has the advantage that it allows for output of the resulting pitch- and note tracks to disk for post-processing. With monophonic music we found it to result pitch tracks that were visually very similar to Melodyne 3. The blue bands in Fig. 3 show the pitch tracks of the neck (top panel) and ankle (bottom panel) recordings, respectively, of the song *Varado* as obtained from the PYIN algorithm in the Tony software. The melody is well-determined from both the neck and the ankle signals (except for a small glitch at the end of the ankle recording).

The still ongoing comparison of recordings from different sensor types at different locations on the body in different experimental setups has so far already demonstrated that body vibration recordings can very well be used for automatic pitch tracking. Although the understandability of the words can be largely reduced in body vibration recordings, since most of the articulation properties of the vocal signal are produced above the larynx, this does not reduce their usefulness for pitch tracking. On the contrary, we encountered several cases in which the pitch tracking performed better on larynx microphone recordings than on regular microphone recordings.

3. ANALYSIS OF POLYPHONIC VOCAL MUSIC

Encouraged by these results, we investigated the usefulness of larynx microphone recordings for automatic pitch analysis of multiple voices in polyphonic vocal music in with two groups of singers. The first group, the trio *La vache qui crie* (Ingrid Hammer, Ursula Häse, Ursula Scribano) performed four songs of different genres and dynamic characteristics (*Akaeli*, a pygmi song; *Heida*, a Georgian krimanchuli (yodel); and *Büchel* and *Summersberger*, two alpine juuz). The recording was done in Traumton studios, Berlin. The results of the analysis of *Heida* are shown in Fig. 4.

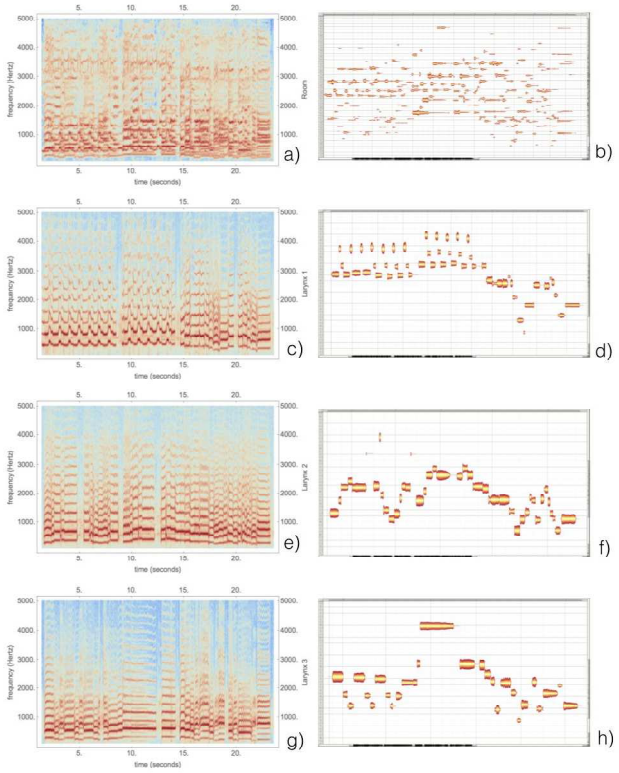


Figure 4. Sonograms (left panels) of the room microphone (a) and the larynx microphone neck recordings of the three singers (b,c,d) singing the song *Heida*. The right panels show the results of the pitch tracking in Melodyne3 to the signals in the left panels.

Despite the wide dynamics of the song and the rapid jumps between chest and head register for singer 1 (c, and d), the melodies of individual voices are cleanly tracked on the larynx microphone recordings in Fig. 4. In contrast, the attempt to pitch track the room microphone recording using Melodyne´s polyphonic algorithm leads to numerous artefacts. The result of using Melodyne´s polyphonic algorithm on room recordings of the other songs (not shown) is essentially the same.

Fig. 5 shows the results of applying the PYIN algorithm (Mauch and Dixon, 2014) to all voices of the complete song *Heida*. Some cross talk between the yodeling top voice (red) and the middle voice (blue) is visible at roughly 18, 19, and 26 in the raw pitch tracks. While this would be easily identified and edited if further processing was desired, here it was left unedited to provide an unbiased documentation of the processing results.

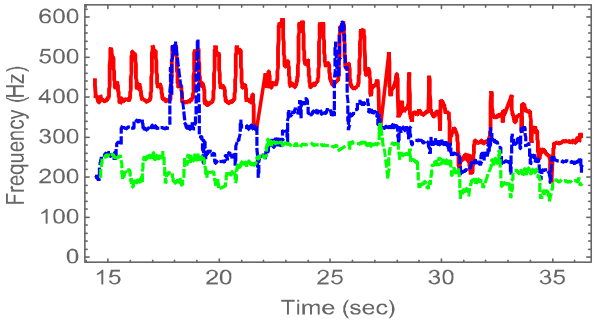


Figure 5. Pitch tracks of the complete song *Heida* as determined by the PYIN algorithm (Mauch and Dixon, 2014).

Fig. 6 shows pitch tracks and spectrograms for the song Büchel.

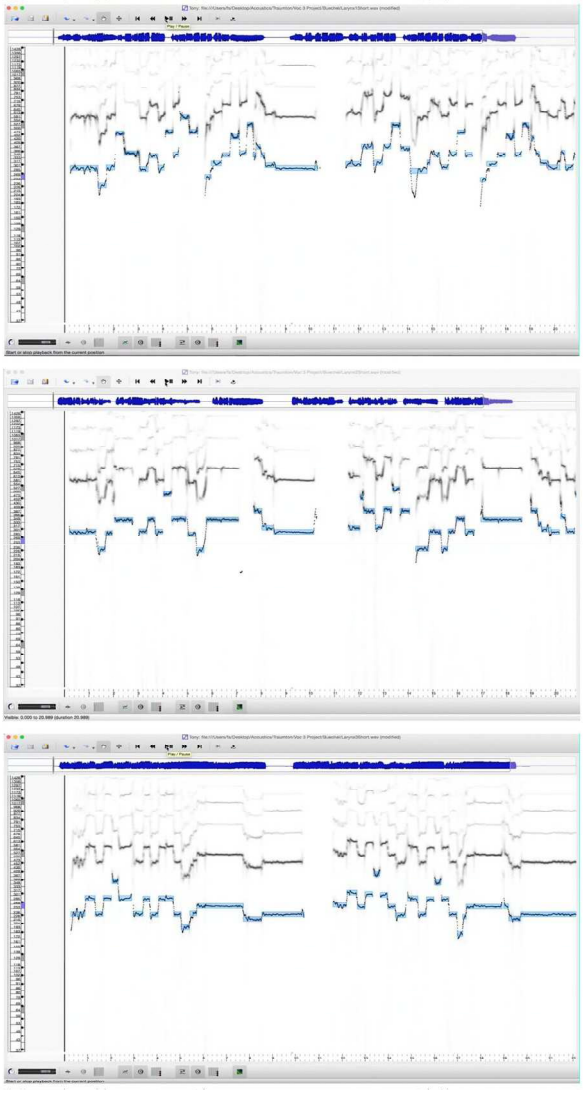


Figure 6. Pitch tracks superimposed on the spectrograms of the complete song *Büchel* as determined by the PYIN algorithm (Mauch and Dixon, 2014).

Again, it can be seen that the melodies of the individual voices are well-identified by the pitch tracks and no obvious cross talk is visible. In comparison, Fig. 7 shows the pitch track of the regular microphone recording of singer 3, which clearly shows problems in identifying the melody (compare with pitch track in lowest panel of Fig. 6).

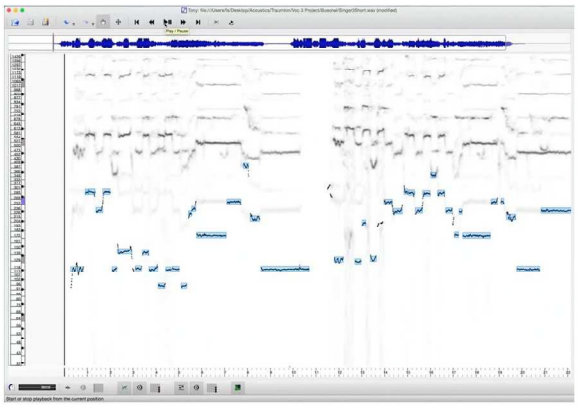


Figure 7. Pitch track superimposed on the spectrogram of the recording of a regular microphone pointed at the mouth of singer 3, performing the song *Büchel*.

To further illustrate this problem, Fig. 8. shows the comparison of the spectrograms (in two different frequency resolutions) and the pitch tracks determined from the larynx microphone mounted at the neck (red curve) and the regular microphone (blue curve). The spectrogram of the larynx microphone (left panel) shows a simple, harmonic spectral structure which may explain the good performance of the pitch tracking algorithm (which is based on analysis of the autocorrelation function). In comparison, the spectrogram of the regular microphone (right panel) shows a more complex structure, the reason for which might be in the stronger contributions from resonances in the vocal tract of the singer but might also stem from cross talk with the sound from the other singers.

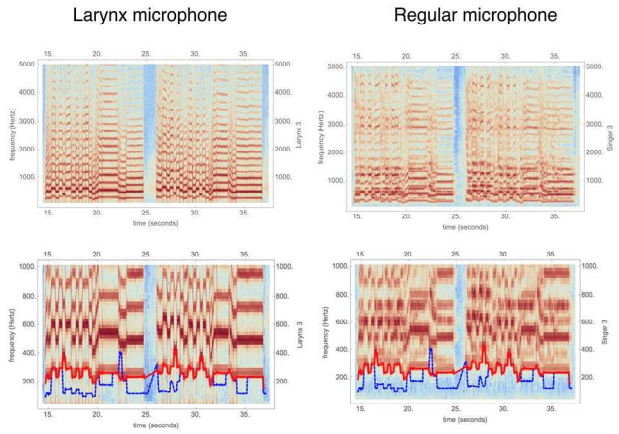


Figure 7. Comparison of the spectrograms of the larynx microphone of singer 3 (left panel) with the spectrogram of the regular microphone. The red and blue curves are the pitch tracks determined from the larynx and the regular microphone, respectively.

As an additional experiment, the second group, the Bulgarian duo (Deniza Popova and Katja Tasheva) performed three diaphonic songs (*Zamraknalo stado*, *Marta Marta*, and *I slanceto treperi*). For this singing style, close interaction between singers is essential. The results for the song *Marta Marta* are shown in Figs. 8 and 9.

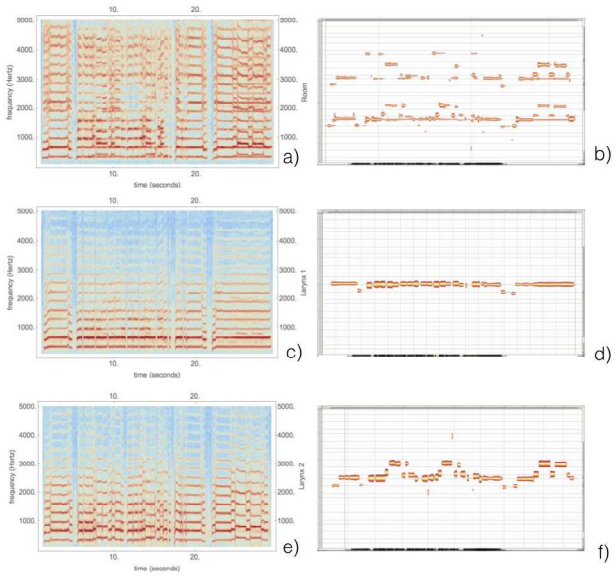


Figure 8. Sonograms (left panels) of the room microphone (a) and the neck-mounted larynx microphone recordings of the two singers (c, d) singing the song *Marta Marta*. The right panels show the results of the pitch tracking.

Again, the individual voices are cleanly pitch tracked on the larynx microphone recordings while the attempt to pitch track the room microphone recording essentially fails.

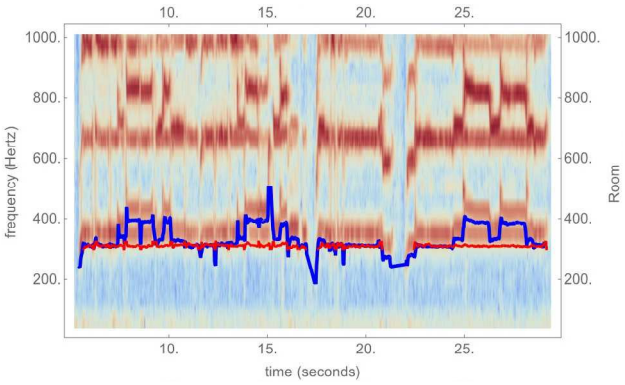


Figure 9. Pitch tracks of the two voices of *Marta Marta* superimposed on the spectrogram of the room microphone.

Fig. 9 illustrates the closeness of the two voices in pitch (mostly within 100 Hz) which demonstrates the potential usefulness of the corresponding records for analysis of the interaction of the singers in terms of generating roughness and/or how beating patterns, for which this kind of music is famous (*interferential diaphonia*), might develop.

4. CONCLUSIONS

The results of the present study suggest that body vibrations can be a valuable source of information for the analysis of polyphonic vocal music. Recordings with larynx or bone microphones (as an add-on to regular acoustic recordings) show excellent separation of the contributions of individual singers and could be quite useful for the documentation of the performance of individual singers, their interaction (as in interferential diaphonia or in studies of entrainment), or for a semi-automatic transcription including any microtonal structures.

5. REFERENCES

Arom, S. (1976), The use of play-back techniques in the study of oral polyphonies, *Ethnomusicology* , 20 (3), 483-519.

Coleman, R. F. (1988). Comparison of microphone and neck-mounted accelerometer monitoring of the performing voice. *Journal of Voice*, *2*(3), 200–205. doi:10.1016/S0892-1997(88)80077-8.

Grupe, G. (2010). Von der Wachswalze zum virtuellen Orchester. *Musicologica Austriaca*, *29*(2010), 35–56.

Henry, P., & Letowski, T. R. (2007). Bone Conduction : Anatomy , Physiology , and Communication. Army Research Laboratory, ARL-TR-4138.

Heracleous, P., Nakajima, Y., Nakajima, Y., Lee, A., Shikano, K. (2004). Non-audible murmur (nam) speech recognition using a stethoscopic nam microphone. In *Proceedings of ICLP* (pp. 1469–1472).

Heracleous, P., Nakajima, Y., & Lee, A. (2004). Audible (normal) speech and inaudible murmur recognition using NAM microphone. In *Proceedings of EU- SIPCO* (pp. 329–332). Retrieved from http://library.naist.jp/dspace/handle/10061/7976\nhttp://library.naist.jp:12180/dspace/handle/10061/7976

Kirikae, J., Sato, T., Oshima, H., & Nomoto, K. (1964). Vibration of the body during phonation of vowels. Rev. de Laryngologie, Otologie, Rhinologie, 85(5-6), 317–345.

Lamarche, A., & Ternström, S. (2008). An exploration of skin acceleration level as a measure of phonatory function in singing. *Journal of Voice : Official Journal of the Voice Foundation*, *22*(1), 10–22. doi:10.1016/j.jvoice.2006.08.005

Matic, A., Osmani, V., & Mayora, O. (2012). Speech activity detection using accelerometer. Annual International Conference of the IEEE Engineering in Medicine and Biology Society., 2112–5.

Mauch, M., & Dixon, S. (2014). PYIN : A fundamental frequency estimator using probabilistic threshold distributions, *1*(1), 659–663.

Mauch, M., Cannam, C., Bittner, R., Fazekas, G., Salamon, J., Dai, J., … Dixon, S. (2015). Computer-aided Melody Note Transcription Using the Tony Software : Accuracy and Efficiency. In *Proceedings of the First International Conference on Technologies for Music Notation and Representation* (p. 8). Retrieved from https://code.soundsoftware.ac.uk/projects/tony/.

Mauch, M., Cannam, C., & Fazekas, G. (2014). Efficient Computer-Aided Pitch Track and Note Estimation for Scientific Applications. In *SEMPRE*. Retrieved from http://code.soundsoftware.ac.uk/projects/tony.

McAngus Todd, N. P. (1993). Vestibular Feedback in Musical Performance : Response to Somatosensory Feedback in Musical Performance (Edited by Sundberg and Verrillo). *Music Perception: An Interdisciplinary Journal*, *10*(3), 379–382.

Melodyne 3, Celemony Software GmbH, Valleystraße 25, 81371 München, Germany.

Milutinović, Z., Mijić, M., & Djurica, S. (1997). Activity of the subglottic voice (“chest resonator”): an echo-tomographic and acoustic study. *European Archives of Oto-Rhino-Laryngology : Official Journal of the European Federation of Oto-Rhino-Laryngological Societies (EUFOS) : Affiliated with the German Society for Oto-Rhino-Laryngology - Head and Neck Surgery*, *254*(6), 292–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9248738

Rahman, T., Adams, A. T., Zhang, M., Cherry, E., Zhou, B., Peng, H., & Choudhury, T. (2014). BodyBeat : A Mobile System for Sensing Non-Speech Body Sounds. In MobySys`14 (pp. 2–13). Bretton Woods, New Hampshire, USA. doi:10.1145/2594368.2594386

Rash, C. E., Russo, M. B., Letowski, T. R., & Schmeisser, E. T. (2009). Helmet-Mounted Displays : Sensation , Perception and Cognition Issues, U.S. Army Aeromedical Research Laboratory

Sundberg, J. (1979). Chest vibrations in singers. *STL-QPSR*, *20*(1), 49–64.

Sundberg, J. (1983). Chest Wall Vibrations in Singers. Journal of Speech, Language, and Hearing Research, 26(3), 329–340.

Sundberg, J. (1992). Phonatory Vibrations in Singers : A Critical Review. *Music Perception: An Interdisciplinary Journal*, *9*(3), 361–381.

Svec, J. G., Titze, I. R., & Popolo, P. S. (2005). Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *The Journal of the Acoustical Society of America*, *117*(3), 1386. doi:10.1121/1.1850074

Takada, O. (2014). Sound Energy Measurement of Singing Voice on upper Parts of the Body : A Research in Classical , Pop , Soul and Musical Theatre Singing. In *ISMA 2014, Le Mans* (pp. 445–451). Le Mans.

Verrillo, R. T. (1992). Vibration Sensation in Humans. *Music Perception: An Interdisciplinary Journal*, *9*(3), 281–302.

von Békésy, G. (1960). Experiments in Hearing (p. 745). McGraw-Hill.

Wikström, B.-O., Kjellberg, A., & Landström, U. (1994). Health effects of long-term occupational exposure to whole-body vibration: A review. International Journal of Industrial Ergonomics, 14(4), 273–292.

Yatani, K., & Truong, K. N. (2012). BodyScope : A Wearable Acoustic Sensor for Activity Recognition. In Ubicomp’12 (pp. 341–350). doi:10.1145/2370216.2370269

ON THE FEASIBILITY OF MARKOV MODEL BASED ANALYSIS OF GEORGIAN VOCAL POLYPHONIC MUSIC

Frank Scherbaum
UP Transfer,
University of Potsdam
fs@geo.uni-potsdam.de

Simha Arom
UMR 7206, CNRS-MNHN, Paris,
simha.arom@gmail.com

Frank Kane
Kane.frank@gmail.com

1. INTRODUCTION

Here we present the first results of a feasibility study regarding the usefulness of Markov models (e. g. Meyn and Tweedie, 2005) for the analysis of Georgian vocal polyphonic music. The present analysis, which can be seen as an attempt to explore new tools for the syntactic analysis of vertical concomitances in music (c.f. Arom & Vallejo, 2008; 2010), is based on the collection of folk songs from Svaneti (NW Georgia) described by Akhobadze (1957). For this purpose, we developed a software package in Mathematica 10 (Wolfram Research, 2014), dedicated to musical score analysis.

2. METHODOLOGICAL FRAMEWORK

In the context of the present study we make the assumption that a song can approximately be treated as a temporal process in which discrete harmonic or melodic states change according to probabilistic rules which are implicitly contained in the song itself. In more technical terms, the assumption made is that the sequence of "chords" (or more precisely vertical conjunctions) in a song can be modeled as a discrete Markov chain of harmonic states (e. g. multi-voice chords of particular duration). In this context, the probability that a particular state changes into another one is assumed to depend only on the current state and a state transition probability matrix T (Markov property), which can be estimated from the musical score by statistical analysis.

3. PROCESSING

In order to determine the Markov model for a Georgian song, the score (cf. Fig. 1) is first converted into digital form (musicXML). Subsequently, the pitches in each voice (in Georgian polyphony usually three) are approximated by piecewise functions of time (Fig. 2), which subsequently are combined into a vector function. This enables the easy algorithmic determination of the margins of harmonic states as times where the time-derivatives of this function are non-zero.

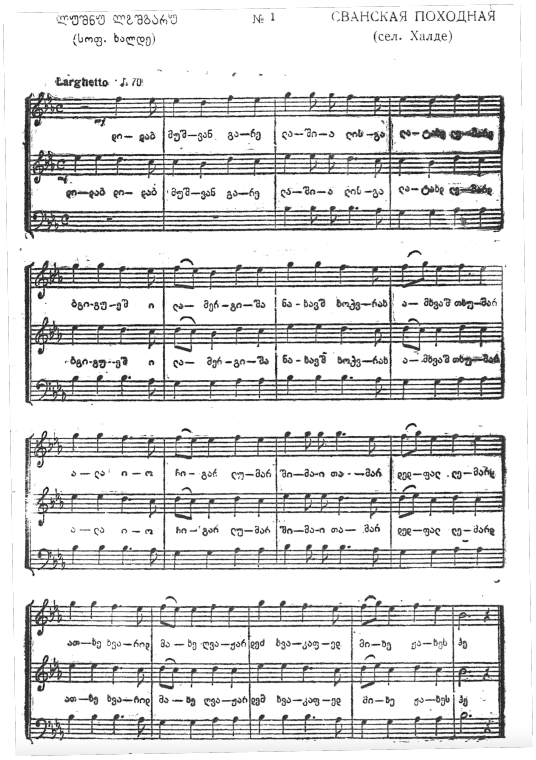


Figure 1. Original score of the song *Lushnu Läshgaru* (Svan marching song).

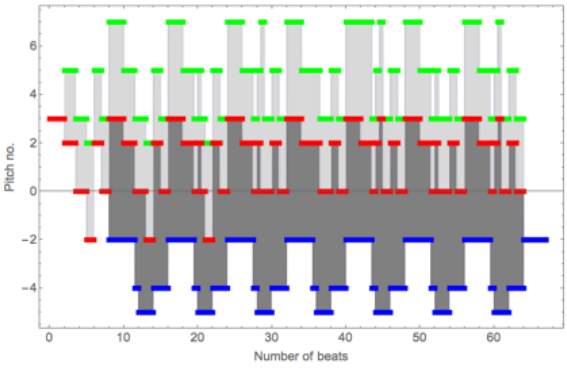


Figure 2. Pitch function of the song *Lushnu Läshgaru*. The red, green and blue lines correspond to the pitches (in pitch number) for the middle, top and bass voice of the song, respectively. These can be seen as components of a vector function, the non-zero derivatives of which define boundaries of the harmonic states of the song.

Since Georgian music is considered modal, we determined the mode of each song (see Figs. 3 and 4) as the mode which is most consistent with the pitch inventory of the song, and converted all chords into mode-degree representations. Regarding the reference note, we followed the strategy of Arom & Vallejo (2008) to use the *finalis*.

Figure 3. Example for the determination of the song´s mode based on the statistical analysis of the pitch inventory of *Lushnu Läshgaru*. The histograms show the distribution of fractional (regarding the duration of the song) consistency (to the right) and inconsistencies (to the left) of the pitch inventories of the individual voices. The center panel shows that the complete pitch inventory of *Lushnu Läshgaru* for all voices is fully consistent with the mode Sol (G), but none other.

Total	Top	Middle	Bass	Mode
0	0	0	0	Sol
0.291667	0	0	0.291667	Do
0.375	0.0333333	0.341667	0	Re
0.841667	0.3	0.25	0.291667	Fa
0.891667	0.316667	0.341667	0.233333	La
1.2	0.316667	0.65	0.233333	Mi
1.51667	0.633333	0.65	0.233333	Si

Figure 4. Mode inconsistency table of the pitch inventory of *Lushnu Läshgaru*. The entries in the different columns indicate the fractional inconsistencies in all voices combined (column 1) and the individual voices (columns 2 – 4) for the tested modes (indicated in column 5). As can be seen from the total inconsistency value of 0, that the complete pitch inventory of *Lushnu Läshgaru* for all voices is fully consistent with the mode Sol (G). For the mode Do (C), there would be inconsistencies with the pitches in the bass melody.

Subsequently, the state transition matrix T (Fig. 5) is determined by statistical analysis of the actually occurring state transitions in the song.

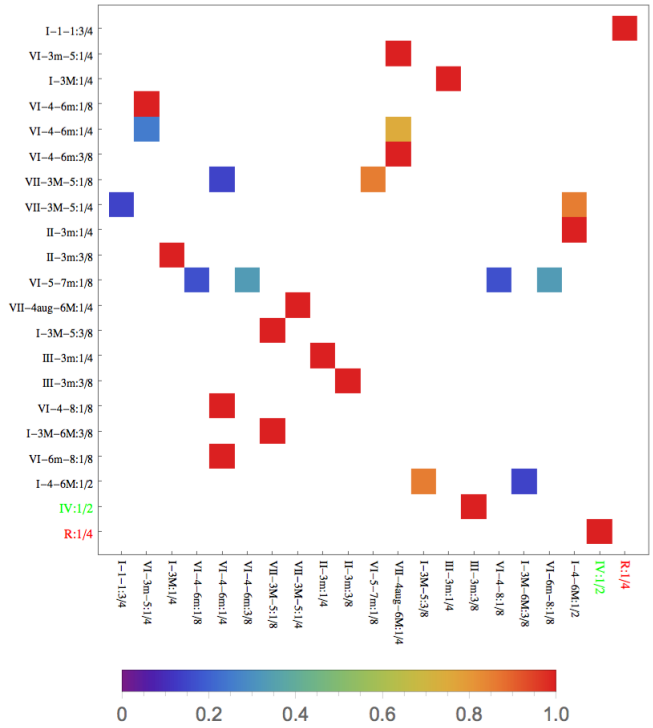


Figure 5. State transition matrix T of the complete song *Lushnu Läshgaru*.

The color code of an element of T describes the relative frequency (loosely speaking the probability) with which the state indicated by the row label is changing into the state indicated by the column label. Elements which are white, correspond to state transitions with zero probability. For example the 3rd row from the bottom corresponds to the state I-4-6M:1/2. It contains two non-zero elements, on for the transition to the state I-3M-5:3/8 with transition probability close to 0.8 (orange color) and one for the transition to the state I-3M-6:3/8 with transition probability close to 0.2 (blue color). Elements with red color describe transitions which take place with probability of 1. In other words, if song comes to the state described by the row label for a red element, the next state is fully determined.

From T we generated the so-called process graphs (Fig. 6) which within the Markov model framework contains the complete information about the corresponding song in a visual (but still quantitative) way. Each node corresponds to a harmonic state, while the edges of the graph represent possible pathways through the harmonic state inventory with their corresponding probabilities color coded as shown in the legend. The start and end states of the song are indicated by green and red text colors, respectively. Process graphs are convenient tools to visually compare songs with respect to their complexity and hence to the predictability of their chord progressions.

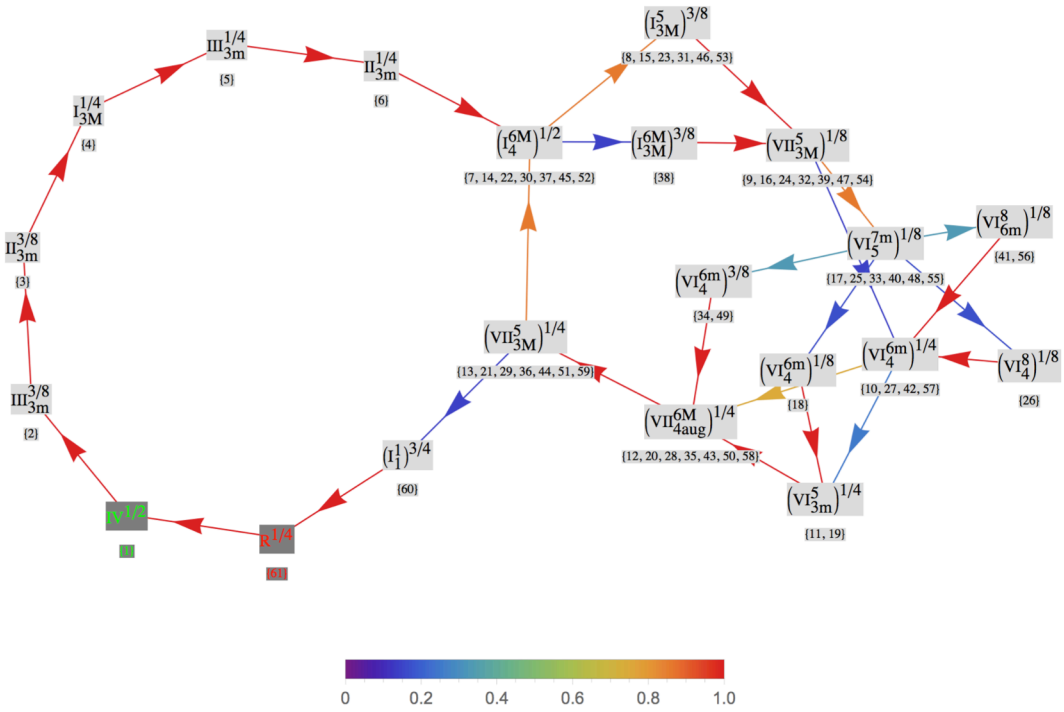


Figure 6. Markov model process graph (right panel) of the complete song *Lushnu Läshgaru*. The chord progression of the actual song (expressed in mode degree notation) can be reconstructed from the sequence indexes below the graph´s nodes.

For the structural analysis of the chord progressions, we removed the introduction, which is typical for many Georgian folk songs (c.f. Fig. 1), and extracted the three-voice part of the song. This results in a considerably simpler process graph (Fig. 7).

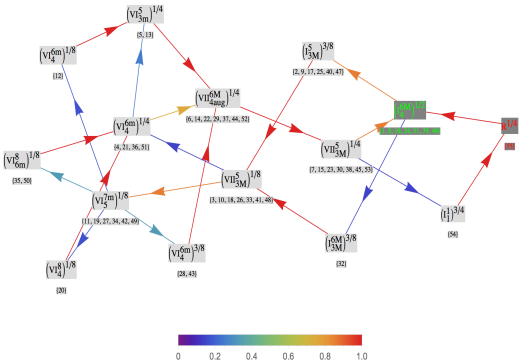


Figure 6. Markov model process graph of the three-voice part of *Lushnu Läshgaru*.

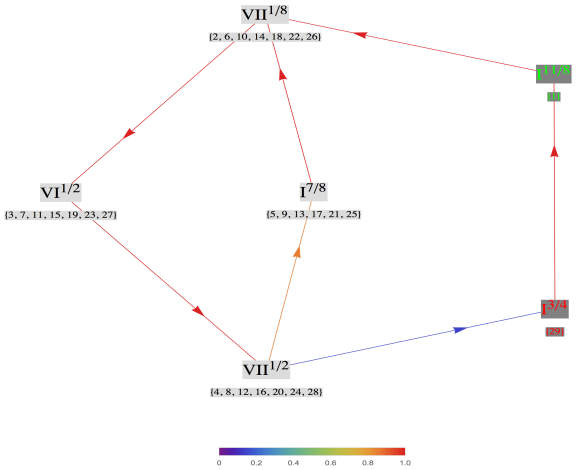


Figure 7. „Chord progression skeleton“ of *Lushnu Läshgaru*.

Subsequently we also removed the chord ciphers to obtain the „chord progression skeleton“ of the song (Fig. 8), which turns out to be a convenient tool for the structural analysis. In the present case for example it shows that the „building block“ of the song consists of the sequence of mode degrees I-VII-VI-VII which is repeated seven times.

In addition to the process graphs, we spot-checked the quality of individual Markov model representations by listening to random realizations (new synthetic songs) of the models generated from the estimated transition matrices. In general, provided the songs are long enough to allow for a robust statistical analysis of states, the Markov model audibly seems to capture the main characteristics of a song in terms of chord progressions rather well.

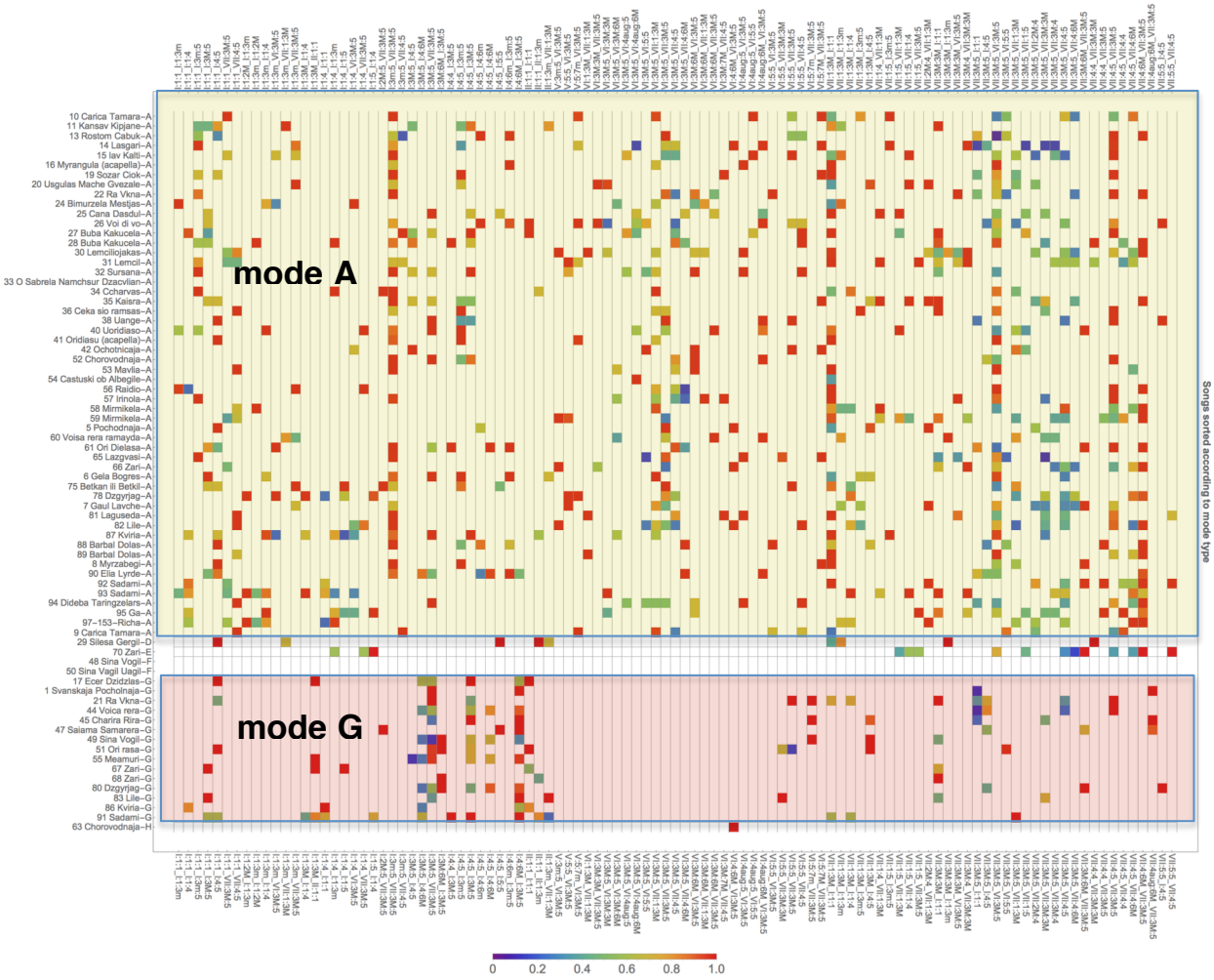


Figure 9. Chord progression profiles regarding the 100 most frequent chord progressions in the whole corpus. The color code of each cell corresponds to the relative frequency with with this chord progression takes place in the corresponding song. Labels are in mode degrees.

4. RESULTS

The state transition matrix T of an individual song (e. g. Fig. 5) can be seen as a simple list of numbers which describe the transition probabilities of all possible chord progressions in that particular song. Usually, the majority of these numbers will be zero because only a small subset of the possible state changes in a song actually takes place. Similarly, for each song one can also determine the transition probabilities of all chord progressions possible in the whole corpus, simply by adding further zeros for all combinations of chords which are in the corpus but not in the particular song. We call this vector the chord-progression vector or -profile of the song. The advantage of doing so is that the chord progression profiles in all songs in the corpus are now of the same dimension and can be quantitatively compared easily. One way of doing so is by projecting all of them into subspaces which are small enough to be visualized (Fig. 9).

Each row in the grid in Fig. 9 corresponds to one song while each column corresponds to one of the 100 most frequent chord progressions in the whole corpus. The songs are sorted vertically according to their modes. The majority of the songs are in mode A (upper yellow shaded rows) and mode G (lower red shaded rows). Those columns which stick out visually as vertical structures correspond to chord-progressions common to all songs of that mode category and are labeled in mode degree notation at the top and the bottom of the panel.

Representing songs in terms of feature vectors (both chord profiles and chord-progression profiles) enables the application of powerful tools from the fields of machine learning and high dimensional visualization to extract information from musical scores. As another example in that direction, Fig. 10 shows the so-called Sammon´s map (Sammon, 1969), which displays the mutual distances of the chord-progression profiles shown in Fig. 9 in such a way that all mutual distances in Fig. 10 are reflecting the mutual distances of the chord-progression profiles in the high-dimensional space.

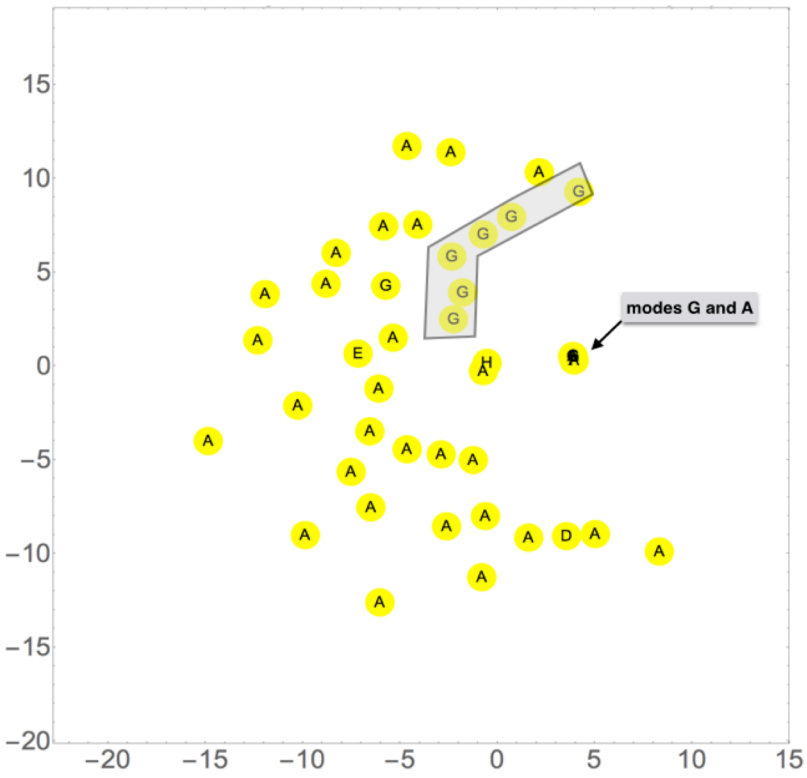


Figure 80. Sammon´s map of the chord-progression profiles of Fig. 9.

Fig. 10 can be directly interpreted as a map of the similarity of the songs in terms of their chord-progressions. Songs which plot close to each other will have more similar chord-progressions than those projecting farther apart. Even without further analysis, Fig. 10 already visually indicates that some of the characteristic chord progressions are mode related.

5. CONCLUSIONS

The present study illustrates different ways to computationally extract information from digital musical scores of polyphonic vocal music and to compare songs quantitatively in terms of chord- and chord-progression structure. The results suggest that Markov models in conjunction with graph theory and high-dimensional visualization techniques provide a powerful and principled framework to perform analysis of musical scores regarding their vertical organization (e. g. Arom and Vallejo, 208; 2010) in semi-automated and completely reproducible ways.

6. ACKNOWLEDGMENT

F.S. is grateful to Giulio Alessandrini from Wolfram Research for fruitful discussions and for sharing his EWTC 2014 presentation on sound processing which stimulated part of the present analysis. We also thank Thoralf Dietrich for his help with the transcription of the scores.

7. REFERENCES

Akhobadze, Vladimer. (1957). *Collection of Georgian (Svan) Folk Songs*. Tbilisi: Shroma da Teknika. (In Georgian. Foreward in Georgian and Russian).

Arom, S., & Vallejo, P. (2008). Towards a theory of the chord syntax of Georgian Polyphony. In Proc. of the 4th International Symposium on Traditional Polyphony, International Research Centre for Traditional Polyphony/ Tbilisi State Conservatory, pp. 321-335, Tbilisi.

Arom, S., & Vallejo, P. (2010). Outline of a syntax of chords in some songs from Samegrelo. In Proc. of the 5th International Symposium on Traditional Polyphony, International Research Centre for Traditional Polyphony/ Tbilisi State Conservatory, pp. 266–277, Tbilisi.

Meyn, S. P. & Tweedie, R. L. (2005). Markov chains and stochastic stability, pp. 567, Springer.

Sammon, J. W. (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, *C-18*(5), 401–409.

Wolfram Research Inc. (2014). Mathematica, Version 10.0, Champaign, IL, USA.

A METHOD FOR TONIC FREQUENCY IDENTIFICATION OF TURKISH MAKAM MUSIC RECORDINGS

Hasan Sercan Atlı
Bahçeşehir Üniversitesi
hsercanatli@gmail.com

Barış Bozkurt
Koç Üniversitesi
barisbozkurt0@gmail.com

Sertan Şentürk
Universitat Pompeu Fabra
sertan.senturk@upf.edu

ABSTRACT

Karar is the final tone in the Turkish makam music performances. The karar frequency, hence the concert pitch varies among performances due to the existence of many diapasons in Turkish makam music, instead of a single standard one. Correct estimation of the karar frequency is critical for many computational tasks such as tuning analysis, audio-score alignment and automatic transcription. We present a new karar frequency identification method that is based on detecting the last note in the recording and estimating its frequency. The method is applied on two large datasets of Turkish makam music recordings and shown to outperform the state-of-the-art.

1. INTRODUCTION

In Turkish Makam Music, the melodic dimension is explained by the *makam* (melodic structure) and the *form*. *Makam* defines the melodic progression from an initial tone to a final tone. Gürmeriç describes the *makam* as "Makam, before everything else, is based on a scale. *Makam* is a progression that gives the *makam* a life by starting from somewhere of the *seyir* (melodic progression), moving towards the *güçlü* (dominant) and going towards *karar*" (Gürmeriç, 1966), (Bozkurt et al., 2014). *Karar* (final) is typically used synonymous to tonic in Turkish makam music. Theoretically and as a very common practice, Turkish makam music performances end at the karar note (Akdoğdu, 1989). [1]

There is no commonly agreed reference frequency (such as A4 = 440 Hz in euro-genetic musics) in Turkish makam music. Moreover the musicians might chose to play the music in one of the 12 *ahenk*s (diapason) due to instrument/vocal range or for musical aesthetics. The tonic note depends on the performed *makam* [2]. The tonic frequency is required for many computational tasks such as automatic transcription (Benetos & Holzapfel, 2013), tuning analysis (Bozkurt, 2012), makam recognition (Gedik & Bozkurt, 2010) and audio-score alignment (Şentürk et al., 2014). While manually annotating the tonic frequency is relatively straightforward (e.g. isolating a tonic note in the performance and adjusting the frequency of a tone which fits this note), it is a time-consuming task. Thus, automatic tonic identification greatly facilitates research based on melody, especially on large data collections.

[1] For the rest of the paper, the note names are written all in lower case and the first letter of a makam name is written in capitals for clarity.

[2] e.g. *karar* of the Rast *makam* is the rast note (G); *karar* of the Hicaz *makam* and Saba *makam* is the dügah note (A).

Pitch distributions (such as *pitch* histograms and *pitch-class* histograms) are commonly used in automatic tonic identification (Gedik & Bozkurt, 2010; Chordia & Şentürk, 2013). In (Krumhansl & Shepard, 1979) 12-dimensional pitch frequency distribution are used to study euro-genetic musics. Pitch distributions are also used for relevant tasks such as key detection and chord recognition (Gómez, 2006) for euro-genetic musics.

For computational analysis of Turkish makam music recordings, the distributions are computed from a melody feature such as fundamental pitch (Gedik & Bozkurt, 2010) or predominant melody (Şentürk et al., 2013). Gedik & Bozkurt (2010) have proposed the so-called *makam histogram template matching* (MHTM) method for automatic tonic identification. In this method the makam of the target audio recording is known. This method computes a template pitch histogram for the *makam*. This template is computed from a set of audio recordings in the same *makam*. Similarly a pitch histogram is computed for the audio recording with unknown tonic frequency. The pitch histogram is shifted and compared against the template histograms. The tonic frequency is identified from the best matching shift. While this method is shown to provide reliable results, it requires training data and *makam* information. Moreover the method might fail in audio recordings with multiple *makam*s (e.g. multiple compositions, improvisations in different *makam*s) and compound makams, which show the characteristics of multiple makams.

Instead of the histogram templates, Şentürk et al. (2013) use machine readable scores to identify the tonic frequency. In this method, performed composition in the audio recording is known and the score for the composition is available. The method extracts the predominant melody from the audio recording. Then kernel-density estimation is applied to the predominant melody to obtain a pitch-class distribution. The peaks of the distribution are selected as tonic candidates. Next the method attempts to partially align the score with the audio recording. The alignment with the best score indicates the tonic frequency. This method provides almost perfect results. However it is not generalizable since it requires composition information and music scores may not be available for all performances (e.g. improvisations).

In this paper we propose a method, which uses the musical knowledge that a makam music performance ends in the tonic note. Our method detects the last note in an au-

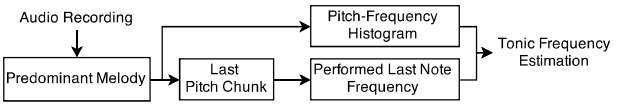


Figure 1: Flow diagram of the LND method

dio recording and identifies the tonic frequency from the pitch estimation of this note. We term this methods as the last note detection (LND) method. This method provides a simpler solution than the current state-of-the-art. Moreover, it does not require any additional information such as the performed *makam* or the composition, hence it is more generalizable. We evaluate the LND method with on two different datasets and show that it outperforms the MHTM method.

The rest of the paper is as follows: Section 2 explains the steps of the MHTM and the LND methods. Section 3 represents experiments and test datasets and Section 4 wraps up the paper with a brief conclusion.

2. METHODOLOGY

The methodology proposed in this paper assumes that Turkish makam music recordings ends on the tonic note. [3] Our method first extracts the predominant melody from the audio recording. The end of the predominant melody is divided into chunks according to jumps on the pitch and the last chunk is estimated as the last note. Initially we estimate the tonic frequency as the median of all the frequencies in this chunk. Next we compute a pitch histogram using the frequency values in the predominant melody, which are close to the initial estimation ($\pm$ a semitone). Then we refine our estimation as the frequency of the closest peak in the histogram. The flow diagram of the LND method is shown in Figure 1.

Below we explain the steps of the methodology in detail.

2.1 Predominant Melody Extraction

Fundamental frequency estimation or predominant melody extraction of an audio recording is the first step of tonic frequency identification methodologies. In Bozkurt (2008), Bozkurt uses YIN (De Cheveigné & Kawahara, 2002) algorithm and with some post-processing steps (e.g. octave correction) to obtain fundamental frequency estimation of monophonic Turkish makam music recordings. Although the YIN algorithm works fine on monophonic recordings, it is not reliable in polyphonic (e.g. ensembles) recordings of Turkish makam music (Şentürk et al., 2014). Bozkurt (2008) also reports that YIN is not reliable in low energy parts (of the signal) with high background noise.

[3] This information is confirmed by the masters we have consulted, such as the kanun virtuoso Reha Sağbaş. Moreover compositions ending with a note other than the *karar* of the relevant *makam* is extremely rare. For example, there is a single example out of 2200 scores in the SymbTr (Karaosmanoğlu, 2012), the largest machine readable score database of Turkish makam music.

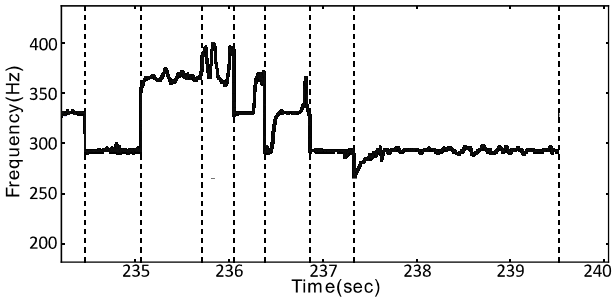


Figure 2: Last 8 pitch chunks of an audio recording. The vertical lines mark the borders of pitch chunks.

Şentürk et al. (2014) use the predominant melody extraction algorithm proposed by Salamon & Gómez (2012) with modified parameters according to the culture-specific aspects of Turkish makam music. It is reported that the algorithm provides better results for both polyphonic and noisy than YIN for polyphonic and noisy recordings. In LND method we use a simplified version of the approach explained in Atlı et al. (2014). This version is observed to provide better pitch estimations especially for the parts of the recording the musicians play comparatively softer.

2.2 Note Boundary Estimation

To observe the last note in LND method, the end of the extracted predominant melody is divided into pitch chunks. The boundaries are placed where the slope of the melody is higher than a specified threshold. In Figure 2, last 8 pitch chunks of a recording [4] is shown.

The last chunk in the predominant melody is estimated as the tonic note. The method computes the median of the frequency values in this chunk as the initial tonic frequency estimation. We prefer median instead of mean since median will be resilient to pitch transitions such as slides and vibratos.

2.3 Pitch Histogram Computation

In LND method, we don't need to compute an overall pitch histogram since we have already obtained an initial estimation. Instead we compute a pitch histogram only from the frequency values in the pitch chunk, which are closer than 100 cents to our initial estimation. We use $\frac{1}{3}$ Holderian comma [5] bin resolution, which is the smallest audible interval in Turkish makam music (Gedik & Bozkurt, 2010).

2.4 Tonic Frequency Refinement

After the pitch histogram is computed for the frequencies around our initial estimation, the peak points of histogram are calculated. We use *pypeaks* [6] library for peak computation. We identify the tonic frequency as the frequency

[4] http://musicbrainz.org/recording/26da8cac-5757-4494-a214-25ad564fc292

[5] 1 Holderian comma is equal to $\frac{1200}{53} = 22.64$ cents

[6] https://github.com/gopalkoduri/pypeaks

Makam(TD1)	#	Form(TD1)	#	Makam(TD2)	#	Form(TD2)	#
Hicaz	135	Şarkı	375	Hicaz	146	Şarkı	412
Nihavend	87	Taksim	131	Hüzzam	95	Taksim	180
K.hicazkar	74	Peşrev	125	K.hicazkar	88	Peşrev	137
Hüzzam	64	Saz Semaisi	92	Nihavend	74	Saz Semaisi	99
Uşşak	62	Ayin-i Şerif	18	Uşşak	67	Ayin-i Şerif	30
Other (25)	446	Other (34)	127	Other (67)	623	Other (38)	235
Total	868		868	Total	1093		1093

Table 1: The makam and form statistics of the TD1 and TD2 datasets

value corresponding to the closest peak to our initial estimation.

3. EXPERIMENTS

To compare against our method, we have re-implemented the MHTM method (Gedik & Bozkurt, 2010). For pitch extraction, the predominant melody estimation methodology explained in (Atlı et al., 2014) is used. *L1-norm* is used as the distance measure for the *recording - template histogram* matching step, which was reported as the best distance measure for this task (Gedik & Bozkurt, 2010).

3.1 Test Datasets

We compiled two test datasets from the Turkish makam music research corpus (Uyar et al., 2014) of CompMusic project (Serra, 2011) [7] . The datasets consist of the audio recordings, the related metadata (such as track names, performers, *makam*s and *forms*) and the tonic annotations.

The first test dataset (*TD1*) is compiled for the comparison of the MHTM method and the LND method. Because of the limitations of the MHTM (Section 1), audio recordings in a single *makam* have been selected. TD1 includes 868 audio recordings in 30 *makam*s which we selected as the most common *makam*s in (Gedik & Bozkurt, 2010). For the evaluation of the LND method, test dataset 2 (*TD2*) is arranged. 225 polyphonic audio recordings where more than one *makam* is performed have been added on TD1 to create TD2. Only the LND method has been tested on this test dataset. TD2 includes 1093 audio recordings, more than 7000 of metadata entries. More than 3000 of these entries are culture-specific information (e.g. *makam*, *form*).

The tonic frequencies of the audio recordings have been manually annotated by the musicians. The annotated frequencies have been matched with the nearest peak point of the pitch histogram of related audio recordings to eliminate the intonation differences. These annotations are used as the ground-truth in the experiments.

Statistics of the test datasets are shown in Table 1. While the commercial recordings cannot be shared (due to the copyright issues), pitch tracks, metadata, and tonic annotations of the audio recordings are available in the Compmusic corpora [8] .

[7] http://compmusic.upf.edu/node/235
[8] http://compmusic.upf.edu/corpora

	TD1			TD2		
Methods	T	F	Success(%)	T	F	Success(%)
MHTM	626	242	72.1	-	-	-
LND	775	93	89.3	964	129	88.1

Table 2: The results of the experiments

3.2 Evaluation and Results

To evaluate the methodologies, we compare the tonic annotations with the estimated tonic frequencies as explained in (Şentürk et al., 2013). If the estimated tonic frequency is within an interval of 1 Holderian comma around the annotated tonic frequency, the estimation is marked as true. Otherwise, the estimation is marked as false. In comparison, the octave differences between the annotated frequency and the estimated frequency are ignored due to ambiguities of the tonic octave in performances, where multiple instruments play the same melody in their own register. The results of experiments are shown in Table 2. [9] T and F refer number of true and number of false estimations, respectively.

3.3 Discussion

The distribution of incorrectly identified tonic frequencies using MHTM method on TD1 is shown in Table 3. It is observed that the tonic frequencies of compound *makam*s have been identified incorrectly. This is expected since the melodic progressions of at least two *makam*s are represented in compound makams. The pitch distributions of the compound *makam* performances are more complicated in comparison to the pitch distributions of the simple *makam* performances. This situation complicates the tonic frequency estimation in audio recordings with compound makams using the MHTM method.

The results obtained from TD1 show that LND method outperforms the MHTM method. Moreover the results obtained from TD2 show that LND method can also be applied to performences where mutliple makams are performed without any drawbacks. Currently, if two notes are smoothly connected with glissando, the LND method cannot divide these notes properly. If this happens in the last two notes, the LND method may identify the tonic frequency as the frequency of penultimate note. As a next step we plan to apply quantisation techniques, which are commonly used in the automatic transcription (Benetos &

[9] The complete results are available at http://compmusic.upf.edu/node/265

Makam	#	F	Err%
Ferahfeza	23	22	95.6
Acemaşiran	12	11	91.6
Hicazkar	30	22	73.3
Sultan-i Yegah	22	15	71.4
Kürdilihicazkar	45	28	60.8
Bestenigar	10	6	60.0
Rast	74	14	31.1

Table 3: The distribution of audio recordings for which the MHTM method fails to identify the tonic correctly in the TD1 dataset

Holzapfel, 2013).

4. CONCLUSION

In this paper, a new methodology for automatic tonic estimation of Turkish makam music recordings is presented. The methodology is based on detecting the last note of a performance (which is also the tonic note) and estimating its frequency. The new method is compared with a previous, more-complicated method. Experiments were performed on two test datasets, which include 1093 audio recordings in total. Our method estimates the tonic frequencies with 89.3% and 88.1% success rate on these datasets. The experiments show that our method outperforms the more-complicated method.

5. ACKNOWLEDGMENTS

This work is supported by the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

6. REFERENCES

Akdoğdu, O. (1989). *Taksim nedir?, Nasıl yapılır?* İhlas A.Ş.

Atlı, H. S., Uyar, B., Şentürk, S., Bozkurt, B., & Serra, X. (2014). Audio feature extraction for exploring Turkish makam music. In *3rd International Conference on Audio Technologies for Music and Media*, Ankara, Turkey. Bilkent University. Faculty of Art, Design and Architecture. Department of Communication and Design.

Benetos, E. & Holzapfel, A. (2013). Automatic transcription of Turkish makam music. In *Proceedings of ISMIR - International Conference on Music Information Retrieval*, (pp. 355–360)., Curitiba, Brazil.

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

Bozkurt, B. (2012). A system for tuning instruments using recorded music instead of theory-based frequency presets. *Computer Music Journal*, *36*(3), 43–56.

Bozkurt, B., Ayangil, R., & Holzapfel, A. (2014). Computational analysis of Turkish makam music: Review of state-of-the-art and challenges. *Journal of New Music Research*, *43*(1), 3–23.

Chordia, P. & Şentürk, S. (2013). Joint recognition of raag and tonic in North Indian music. *Computer Music Journal*, *37*(3).

De Cheveigné, A. & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of 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. (2006). *Tonal Description of Music Audio Signals*. PhD thesis, Universitat Pompeu Fabra.

Gürmeriç, A. (1966). Lecture notes. İBK Türk Müziği Bölümü Talebe Cemiyeti Neşr'yatı.

Karaosmanoğlu, K. (2012). A Turkish makam music symbolic database for music information retrieval: SymbTr. In *Proceedings of 13th International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 223–228)., Porto, Portugal.

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*(4), 579–594.

Salamon, J. & Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(6), 1759–1770.

Şentürk, S., Gulati, S., & Serra, X. (2013). Score informed tonic identification for makam music of Turkey. In *Proceedings of 14th International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 175–180)., Curitiba, Brazil.

Şentürk, S., Holzapfel, A., & Serra, X. (2014). Linking scores and audio recordings in makam music of Turkey. *Journal of New Music Research*, *43*(1), 34–52.

Serra, X. (2011). A multicultural approach in music information research. In *Proceedings of 12th International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 151–156)., Miami, Florida (USA).

Uyar, B., Atlı, H. S., Şentürk, S., Bozkurt, B., & Serra, X. (2014). A corpus for computational research of Turkish makam music. In *1st International Digital Libraries for Musicology Workshop*, (pp. 1–7)., London, UK.

Web analysis tools and ethnomusicology

Joséphine Simonnot*[1], Guillaume Pellerin*[†] , and Thomas Fillon*[‡]

[1]CNRS – CREM-LESC – France

Abstract

For many years, one of CNRS's objectives is to improve access and to facilitate data sharing to the entire academic community, through a toolbox of open source softwares (HumaNum service). In anthropology, ethnomusicology and linguistics, researchers work on multiple kinds of documents such as sound recordings. The need to preserve and to easily access, visualize and annotate these materials is problematic because their diverse formats, contents and the increasing quantity of data. With new audio technologies, questions linked to the preservation, the archiving and the availability of these audio materials have arisen. Since 2007, French ethnomusicologists and engineers have joined their efforts to develop a collaborative web platform for managing and improve access to digitized sound archives. This web platform, an open-source web audio CMS dedicated to digital sound archives (Telemeta), is developed through the expertise of the Parisson Company. Its architecture is associated with TimeSide, an open-source audio processing framework written in Python and JavaScript languages, which provides decoding, encoding and streaming capabilities together with an embeddable HTML audio player. Consequently, this CMS is able to can produce various visualization, annotation, segmentation, etc. Since 2013, automatic musicological analysis tools have been also developed in a national interdisciplinary research project called DIADEMS (Description, Indexation, Access to Documents of EthnoMusicology and Sound). Furthermore it includes a set of audio analysis plug-ins and wraps several audio features extraction libraries. These technologies are shared by a large developer community in audio processing: Queen Mary University, IRCAM, Barcelona University, New York University of Abu Dhabi, etc.
The benefits of this collaborative platform for humanities and social sciences research apply to numerous aspects of the field of ethnomusicology, ranging from musical analysis to comparative history and anthropology of music, as well as to the fields of anthropology, linguistics and acoustics. Some of these benefits have been mentioned in several acoustic and ethnomusicological publications.

This paper introduces how cutting-edge tools are being implemented to fit new ways to access and indexing sound libraries. New needs and new expectations from users will be presented in this paper.

Authors: Guillaume Pellerin (Parisson), Joséphine Simonnot (CNRS, France), Thomas Fillon (Parisson)

*Speaker
[†]Corresponding author: pellerin@parisson.com
[‡]Corresponding author: thomas@parisson.com

SINGER DIARIZATION: APPLICATION TO ETHNOMUSICOLOGICAL RECORDINGS

Marwa Thlithi
IRIT, Univ. Toulouse
118, route de Narbonne
31062 Toulouse – France
thlithi@irit.fr

Claude Barras
LIMSI-CNRS, Univ.
Paris-Sud, 91403,
Orsay, France
Claude.Barras@limsi.fr

Julien Pinquier, Thomas Pellegrini
IRIT, Univ. Toulouse
118, route de Narbonne
31062 Toulouse – France
{pinquier,pellegrini}@irit.fr

1. INTRODUCTION

A music audio document can be structured automatically by many ways according to the final objective. In the context of a project on indexing ethno-musicological audio documents, we asked ourselves the questions: who is singing and when. By analogy with *speaker diarization* which consists in detecting who is speaking and when, we called the fact of detecting changes of singers, *singer diarization*. Figure 1 illustrates the task. The ground truth consists of a manual annotation in singing turns, and eventual entry/exit of instruments.

In the context of the ANR DIADEMS[1] project (*Description, Indexing, Access to ethno-musicological and Sound Documents*) on indexing ethno-musicological audio documents, singer diarization automatically appeared to be essential. In this paper, we present our developed singer diarization system which is applied on ethno-musicological recordings.

The paper is organized as follows. In the next section, we describe our singer diarization system. In section 3, the application context is presented. In section 4, performance is presented and discussed.

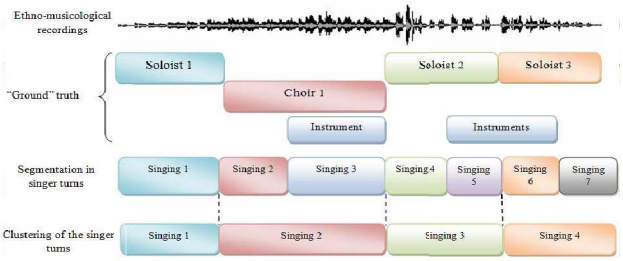


Figure 1. Illustration of singer diarization task.

2. SINGER DIARIZATION METHOD

Singer diarization consists in segmenting musical recordings, and then in labeling segments known as "acoustically homogeneous", our final goal is to obtain segments comprised of the singing of a single group of singers. Singer diarization is divided into two steps: segmentation step and clustering step.

Segmentation step consists in segmenting musical recordings into segments "acoustically homogeneous". Our method for the segmentation step is based on the Bayesian Information Criterion (BIC), which is widely used in audio segmentation (Chen, 1998; Delacourt, 2000; Siu, 1991). Its application on ethno-musicological recordings required an adaptation of two parameters: the size of the signal window, in which a border of segment is searched, and the penalty factor. The adaptation of the window analysis size was solved by implementing a version of the algorithm in which the window size increases while no potential boundary is found (Cettolo, 2005). For the penalty parameter, we observed that no single value was optimal for all the recordings. This led us to propose the Consolidated *A posteriori* Decision (DCAP) method, which consists in combining several segmentations obtained by varying the value of this parameter within the interval [0.8 1.2] with a step of 0.01 (Thlithi, 2014). Then, a vote is carried out on the candidates obtained from all these 41 segmentations: a boundary is validated if it was found by at least S_0 segmentations among all the segmentations. S_0 is determined on a development set. A tolerance gap of 0.5 s was used for this purpose. We used Mel Frequency Cepstral Coefficients (MFCC) for the parameterization step.

In order to achieve the goal of labeling all segments produced by the same group of singers with a unique identifier, an agglomerative clustering is performed on the output of the segmentation step. The signal is also parameterized with MFCC for this clustering step and clusters are modeled by a single Gaussian with a full covariance matrix. In the first step, each segment seeds one cluster. The two nearest clusters according to the BIC criterion are then merged until the criterion sign changes.

3. CORPUS

The "*DIADEMS*" corpus was provided by the ethnomusicologist partners of the DIADEMS project. Examples are accessible online[2]. It is comprised of music recordings with a variable sound quality (outdoors in general, presence of background noise and audio events other than music). These records were done in several sub-Saharian countries. They mainly contain singer turns solo / choir with instruments or speech. This corpus contains 9 music recordings of 20 minutes in total which we divided into a development corpus (DEV) and an evaluation corpus (EVAL) in the proportions 20% and 80%.

1 http://www.irit.fr/recherches/SAMOVA/DIADEMS/en/welcome/

2 http://diadems.telemeta.org/archives/fonds/CNRSMH_DIADEMS/

This corpus was manually annotated in terms of singer turns. A segment boundary is inserted in the following situations:

- Change from a group of N singers G_i (i=1..N) to another group of N' singers G_j' (j=1..N'), such as:

$$G_i \neq G_j' \text{ and } \forall \ i, j$$

- Change from a group of singers to a no-singing area (silence, instruments, speech, etc.) and vice versa.

Then, the segments which contain the same group of singers are annotated with the same label.

4. RESULTS

We used the DEV subset to determine the S_0 parameter of the DCAP method for the segmentation step. We obtained S_0 equal to 15. A 61.2% F-measure was obtained for the segmentation step for the EVAL corpus.

Global results of singer diarization system on the DEV and EVAL corpus are presented in Table 1. Performance of the clustering result is expressed in term of diarization error rate (DER) which is the standard for speaker diarization and measures the fraction of time that is not attributed to the correct label, given an optimum mapping between the reference identifiers and the true labels. The value of the penalty factor for clustering set on the DEV is equal to 6.

Corpus	DER
DEV	17.8%
EVAL	43.1%

Table 1. Global performance for DEV and EVAL corpus.

We noticed that performance varies significantly from one recording to another. Indeed, ethno-musicological recordings are very heterogeneous. Listening to the recordings where we have many errors reveals the presence of superimposed singers, rapid alternations between soloists and a choir, the presence of percussive instruments such as bells, hand claps, and background noise. Moreover, these recordings proved to be more difficult to manually annotate in general.

5. CONCLUSION

In this article, we presented the singer diarization task. The long-term objective is indexing the content of ethno-musicological recordings. For the segmentation step, we applied a method based on the BIC criterion. The choice of optimal single value for the penalty parameter of this criterion proved unsatisfactory. In order to avoid selecting a single value, we combined obtained segmentations with different values, and the final segmentation is obtained by keeping only the boundaries present in several of them. For the clustering step, we applied also the BIC criterion with an adjustment of its penalty parameter on the DEV set.

In order to improve our system, other clustering approaches are currently being tested such as the method of the ILP clustering with i-vectors.

6. ACKNOWLEDGEMENTS

This work was partly funded by the French National Agency for Research (ANR) under grant ANR-12-CORD- 0022-05 (project DIADEMS).

7. REFERENCES

Cettolo, M., Vescovi, M., & Rizzi, R. (2005). Evaluation of BIC-based algorithms for audio segmentation. In *Computer Speech And Language*, (pp. 147-170).

Chen, S.S, & Gopalakrishnan, P.S, (1998). Speaker, Environment and Channel Change Detection and Clustering via the Bayesian Information Criterion. *The DARPA Broadcast News Transcription and Understanding Workshop*.

Delacourt, P., & Wellekens, C., (2000). DISTBIC: a speaker-based segmentation for audio data indexing. In *Speech Communication*, vol. 32, (pp. 111-126).

Siu, M.-H., Yu, G., & Gish, H. (1991). Segregation of speakers for speech recognition and speaker identification. In *Proceeding of International Conference on Acoustics, Speech, and Signal Processing*, Toronto, Canada, (pp. 873-876).

Thlithi, M., Pellegrini, T., Pinquier, J., & André-Obrecht, R., (2014). Segmentation in singer turns with the Bayesian Information Criterion. In *Proceedings of International Speech Communication Association*, Singapore, (pp. 1988-1992).

AN INTERACTIVE RHYTHM TRAINING TOOL FOR USULS OF TURKISH MAKAM MUSIC

Burak Uyar
Bahçeşehir University
burak.uyar@stu.bahcesehir.edu.tr

Barış Bozkurt
Koç University
barisbozkurt0@gmail.com

ABSTRACT

The education of the Turkish makam music practice is carried out by face-to-face sessions with the masters. Our aim in this work is to present a rhythm training software that helps users imitate recording of a master and receive visual feedback about their performances compared to the recording of the master. The system automatically detects onsets in the recordings and then performs comparison of the onset locations with a novel approach defined here. We present a demonstration of the software and preliminary test results on basic rhythmic patterns of Turkish makam music.

1. INTRODUCTION

In Turkish makam music culture, education process is mainly based on practicing the melodic and rhythmic patterns. Being an oral tradition, the fundamentals of this music culture is generally learned from a master in face-to-face sessions. This education model is called *meşk* in Turkish makam music tradition. During *meşk*, the master explains and demonstrates the pattern and then the student practices until the master approves the student's proficiency. This study targets the presentation of a software tool developed to help students practice rhythmic patterns at home.

In this study, we specifically focus on practicing the rhythmic patterns in makam music. The rhythmic patterns are called *usuls* in Turkish makam music. A sufficient description for *usuls* is as stated in Bozkurt et al. [2014], "An usul is a rhythmic pattern of a certain length that contains a sequence of strokes with varying accents." There are a few studies and commercial applications related to *usul* training. One of them is Mus2Okur, which includes a large amount of *usuls* with auditory material and symbolic representations. Another application is Usul-Velvele-Editor, which also provides auditory and visual data for *usuls*. These applications include comprehensive collections but lack interactivity with the user.

Various interactive tutoring systems have been developed previously for other music cultures. An interactive music training system is described by Percival [2008]. In the rhythm training part of the system, the user's performance recording is analyzed and an accuracy score is provided. One other interactive system is developed in the *i-maestro* [1] project. One of the key outcomes of that system is denoted as creating tools for instrumental training in different environments. This system includes both auditory and visual analysis of a user's performance. Wang and Lai [2011] have developed a mobile digital game based tool for rhythm learning. Ferguson [2006] described an interactive real-time system for instrumental practice and tuition. Existence of such tools encourage us to create similar systems for *usul* training.

[1] http://www.i-maestro.org/

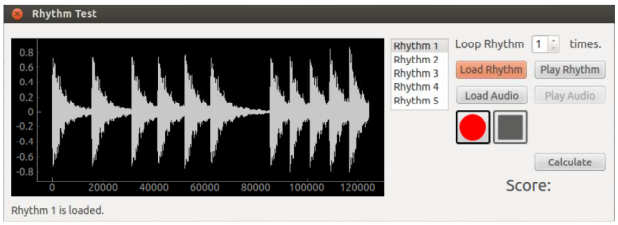


Figure 1: Software interface with a reference rhythm loaded.

In our study we consider the points mentioned above to create a tool for learning and practicing *usuls*. By taking the advantages of the MIR methodologies, it is possible to develop a complementary tool to support the *meşk* process. In case of difficulty in accessing to masters of Turkish makam music, this tool can also be used for self-tutoring. The following parts of the paper include the description of the system with its constraints, data, modules and methods.

2. SYSTEM DESCRIPTION

The main purpose of the program is to provide a tool for practicing *usuls* in an interactive manner. In this system, users can select *usuls* from a list and record their performances after listening the *usul* patterns. After recording, users can use the tool to compare their rhythm performance with the *usul* template. In line with this purpose, the software has an interface to provide simple tools to the user as seen in Figure 1. In the following parts, the constraints of the system, the data used for the current version, modules of the software and the methods are described in detail. Additionally, some features of the system are compared with a system designed for a similar task.

2.1 Constraints

The system is designed in consideration with some environmental and hardware-related constraints. First, the noise in the room should be at a reasonable level so that the onset detection algorithm can provide reliable results. Secondly, the user should be able to interact with the computer by using the speakers and the microphone.

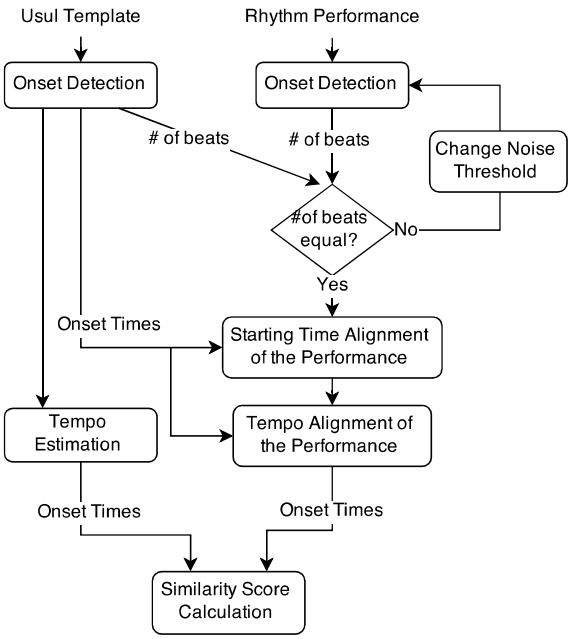


Figure 2: Workflow diagram of the tool.

2.2 Data

The system includes an audio file collection of *usul* patterns with symbolic representations as well as recordings from masters. These *usuls* are obtained from the Turkish makam music corpus, Uyar et al. [2014], considering their prevalence. Then, the most common *usuls* with different time signatures have been selected for preliminary tests. After creating a list of *usuls*, their audio files and symbolic representations have been prepared. The audio files are in Wav format and the symbolic data is prepared as MusicXML [2] scores. While generating the images from MusicXML files, Lilypond [3] is utilized because of its ease of use and compatibility with MusicXML files.

2.3 Modules and Methods

Mainly, there are two work branches in the system. One is for processing the reference rhythm and the other one is for analyzing the performance recording. The information retrieved from the reference rhythm is used for scoring the performance. The general view of the workflow is as seen in Figure 2. For audio signal processing and analysis, the Essentia [4] library is used. The tool is implemented in Python [5] programming language and the current version of the software runs on Ubuntu OS [6]. The system consists of several modules. These are the audio recorder/player, the audio signal visualizer and the audio processor modules.

The audio recorder/player module is implemented by using the PyAudio [7] library. This module enables the users to load a certain *usul* to practice, record a performance or load a file from the file system to calculate a score for it by taking the selected *usul* as the reference. Users can listen to the selected *usul* pattern, their own performances or the file loaded from the file system.

The audio signal visualizer part of the system is implemented by using the library provided by PyQtGraph [8]. By using the plotting functions of this library, the waveform of the reference *usul* is displayed on the interface. After the scoring of the performance or the file loaded from the file system, the onsets are displayed by vertical lines on the reference waveform.

The audio processing module is practically the core of the software. In this module there are functions for detecting onsets in an audio file, creating a binary vector as the onset signal, aligning the onsets detected from two audio files in terms of tempo and starting time, tempo estimation and score calculation processes.

For the onset detection, the *OnsetDetection* [9] function is used as the base with the parameter of *method="flux"*. This parameter sets the function to use the *Spectral Flux* detection method. The default values are used for the frame parameters as *frameSize=1024* and *hopSize=512*, as suggested in the documentation. In this study, the onsets detected by the *OnsetDetection* algorithm are selected by an automatic noise threshold setting. This additional feature makes us to be able to use the system in different environments with different noise levels. To do this, the algorithm changes the *noisethreshold* parameter at every iteration and runs until the number of onsets obtained from the performance recording is the same as the number of onsets in the reference recording.

Since the performance usually does not start as soon as the recording starts, it is necessary to align the starting times of the onsets of the performance and the reference. To accomplish this, a function is implemented to move the detected onset times of the performance providing that the first onset times of the reference and the performance are the same. In this procedure, the first onset time of the performance is subtracted from all onset times of the performance. Then, the first onset time of the reference recording is added to all onset times of the performance. For the tempo alignment, the tempo of the performance is assumed to be constant in order to achieve a consistent rhythmic pattern. With this idea in mind, the tempo alignment algorithm stretches or squeezes the onsets detected from the performance recording in a linear manner. The algorithm compares the durations of the reference and the performance recordings. Then it resizes the performance onset times by a constant ratio of D_R/D_P, where D_R and D_P denote the duration of the reference and the duration of the performance, respectively. The tempo estimation function [10] is used as it is provided in the Essentia library.

For the score calculation, Percival [2008] uses Mean-

[2] http://www.musicxml.com/
[3] http://www.lilypond.org/
[4] http://essentia.upf.edu
[5] https://www.python.org/
[6] http://www.ubuntu.com/
[7] https://github.com/bastibe/PyAudio
[8] http://pyqtgraph.org/
[9] http://essentia.upf.edu/documentation/reference/std_OnsetDetection.html
[10] http://essentia.upf.edu/documentation/reference/std_RhythmExtractor2013.html

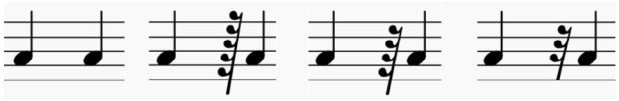


Figure 3: The beats for the perceptional test.

Squared-Error (MSE) to measure the correctness of a performance according to a reference rhythm. After the onsets of the performance are detected and aligned with the reference, the error for each onset is calculated in terms of frames. These errors are used to calculate the MSE. After the MSE is calculated, it is scaled with a constant value which is set after experiments, and the final score is provided by $(1 - MSE)$. However, in our study the criteria were decided to have constraints dependent on musical perception. For this purpose, a simple experiment was executed to understand how the short-timed notes are perceived at different tempo, ranging from 80 beats per minute (bpm) to 120. 5 subjects who are moderately experienced musicians were asked to differentiate the beats in Figure 3 by listening. Our initial listening tests show that the beat without rest and the beat with 128th rest were almost indistinguishable, the beat with 64th rest was recognizable around 80 bpm and the beat with 32nd rest was always recognized. The current scoring algorithm is based on this data. However, more detailed listening tests will be conducted in order to improve the reliability of the scoring algorithm. By considering the results of the initial listening tests, the scores for the different time intervals in terms of note lengths for individual onsets are set as in Table 1. In the scoring procedure, each onset of the performance recording is checked with the corresponding onset of the reference. The time difference between these onsets is calculated. Smaller time differences indicate more accurate performance timings. For example, if the time difference for an individual onset is less than the duration of a 128th note, it is scored as 100% correct. If it is between the duration of a 128th note and a 64th note, the onset is scored as 90% correct. After scoring the individual onsets, the average of the scores of all onsets of the performance except the first and the last ones is calculated to provide the overall score of the performance. An example is provided in Figure 6.

By using the functions described above, the procedure for calculating the score of a performance is as follows. First, an *usul* is selected from the rhythms list. Right after an *usul* is selected, the onsets and the tempo of the selection are calculated in the back-end and the waveform is displayed on the interface. Then, the user listens the *usul* and records a performance. After recording, user can listen the performance or directly click the *Calculate* button. When the button is clicked, the onset detection algorithm runs on the performance recording. The function takes the number of onsets of the reference as a parameter N_R and repeats the analysis by changing the *noisethreshold* until the number of detected onsets is equal to N_R. If the algorithm cannot find onsets providing this condition, it warns the user by a text message and prompts to start again. After finding the onsets, the onset times of the performance and the reference are aligned to start at the same time. Then, the tempo of the performance is aligned with the reference. Once these operations are done, the score is calculated and presented to the user as the correctness score of the performance.

Interval	$\leq$128th	64th	32nd	16th	8th	$\geq$4th
Score	1	0.9	0.8	0.3	0.2	0.1

Table 1: Scores for individual onsets according to intervals indicating the difference from corresponding onset of the reference.

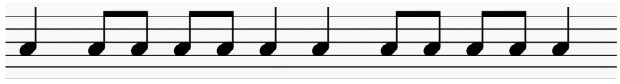


Figure 4: The beat pattern used for comparing our system's output with the Percival's.

2.4 Review of the System

The most similar system provided for measuring the similarity between two rhythm patterns is the one provided in Percival [2008]. Therefore, we will compare our methodology with this one and discuss the pros and cons after summarizing his methodology briefly. In his method, the onsets are detected by calculating the root mean square (RMS) for each frame and comparing the RMS with a preset threshold. The frames with RMS higher than the threshold are considered as onsets. Once the onsets are detected and the start times of the performance is aligned with the reference, the MSE is calculated to provide the score of the performance. The first difference between this system and our system is the onset detection method. While the threshold for detecting the onset frames is constant in Percival's method, using the *spectral flux density* in our method enables users to use the system in rooms with different stable noise levels. In Percival's system, rhythm patterns in different tempos are treated by the same timing parameters. In the system we describe, the scores of the onsets are defined in terms of musical note lengths, which are dependent on the tempo. This provides measuring the correctness of a performance according to the perceptional criteria. Another difference is that in our methodology, the tempo of the user's performance does not affect scoring. This is because we mainly focus on learning the rhythm pattern, not playing it at the same tempo as the reference recording. To provide a basic comparison, the reference beat pattern (Figure 4) and the performance used to test Percival's method are regenerated. The data can be reached at GitHub [11] . For this data, Percival's method results a final grade of 68% (Figure 5), where our method provides an accuracy score of 85% (Figure 6).

[11] https://github.com/burakuyar/rhythm_tool_test_dataset

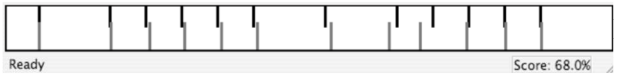


Figure 5: Output of Percival's system for the test data.

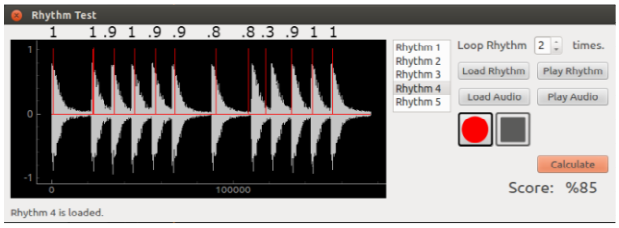


Figure 6: Output of our system for the test data.

Currently, the implementation of the system is completed. However, it is tested for limited number of scenarios. According to the experiments conducted until now, it is observed that the system can distinguish very good and very bad performances. As the next step, 10 different *usuls* will be selected and the performances of these *usuls* will be recorded from 5 different users. These performances will be scored by music experts and will be ordered considering these scores. Then, our algorithm will be optimized in order to provide the same ordering.

3. CONCLUSION AND FUTURE EXPECTATIONS

In this paper, an interactive tool for learning and practicing rhythmic structures, *usuls*, of Turkish makam music is described. In parallel to the methodology, a collection of symbolic beat scores and audio recordings describing the basic *usuls* are prepared. The tool presented analyzes two different recordings of a rhythmic pattern and presents an accuracy score with respect to the preset conditions. The current method for score calculation will be improved by conducting experiments and comparing the outputs with the outputs of the method provided by Percival [2008]. We hope that this study would stimulate research on novel interfaces and computational methodologies for the analysis and education of Turkish makam music.

4. ACKNOWLEDGMENTS

This work is partly supported by the European Research Council under the European Unions Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

5. REFERENCES

B. Bozkurt, R. Ayangil, and A. Holzapfel. Computational analysis of turkish makam music: Review of state-of-the-art and challenges. *Journal of New Music Research*, 43(1): 3–23, 2014.

Sam Ferguson. Learning musical instrument skills through interactive sonification. In *Proceedings of the 2006 conference on New interfaces for musical expression*, pages 384–389. IRCAMCentre Pompidou, 2006.

Mus2Okur. Multimedia encyclopedia of turkish music. URL `http://www.musiki.org/Mus2okur_en.aspx`.

G.K. Percival. Computer-assisted musical instrument tutoring with targeted exercises, university of victoria. 2008.

Usul-Velvele-Editor. Usul-velvele editor. URL `http://notist.org/usul_velvele_editoru.html`.

B. Uyar, H.S. Atlı, S. Şentürk, B. Bozkurt, and X. Serra. A corpus for computational research of turkish makam music. In *1st International Digital Libraries for Musicology Workshop, London*, pages 57–65, 2014.

Ching-Yu Wang and Ah-Fur Lai. Development of a mobile rhythm learning system based on digital game-based learning companion. In *Edutainment Technologies. Educational Games and Virtual Reality/Augmented Reality Applications*, pages 92–100. Springer, 2011.

MULTILEVEL MELODIC MATCHING

Chris Walshaw
Department of Computing & Information Systems,
University of Greenwich, London SE10 9LS, UK
c.walshaw@gre.ac.uk

ABSTRACT

This paper describes a multilevel algorithm for matching tunes when performing inexact searches in symbolic musical data. The basis of the algorithm is straightforward: initially each tune in the search database is normalised and quantised and then recursively coarsened, typically by removing weaker off-beats, until the tune is reduced to a skeleton representation with just one note per bar. The same process is applied to the search query and melodic matching between query and data can then take place at every level. The algorithm implemented here uses the longest common substring algorithm at each level, but in principle a variety of similarity measures could be used. The multilevel framework allows inexact matches to occur by identifying similarities at course levels and is also exploited with the use of early termination heuristics at coarser levels, both to reduce computational complexity and to enhance the matching qualitatively. Experimentation demonstrates the effectiveness of the approach for inexact melodic searches within a corpus of tunes.

1. INTRODUCTION

This paper presents a multilevel melodic matching algorithm. In a sister paper a variant is explored as a method for identifying related tunes (Walshaw, 2015) but here it is used to perform inexact melodic searches, in particular within the abc notation music corpus.

Abc notation is a text-based music notation system popular for transcribing, publishing and sharing music, particularly online. Similar systems have been around for a long time but abc notation was formalised and named by the author in 1993. Since its inception he has maintained a website, now at abcnotation.com, with links to resources such as tutorials, software and tune collections.

In 2009 the functionality of the site was significantly enhanced with an online tune search engine which indexes a corpus of around 460,000 abc transcriptions from across the web. Users of the tune search are able to view, listen to and download the staff notation, MusicXML, MIDI representation and abc code for each tune, and the site currently attracts around half a million visitors a year.

Currently, however, the search engine is purely text based: it is possible to search melodically, since abc is a text based notation system, but only for exact text matches, and as a result most of the searches that take place are based on tune titles (and other meta data).

The motivation for this paper is to propose an inexact melodic search which will identify tunes closely related to the search query, an important feature for an index of what is mostly folk and traditional music.

2. RATIONALE: A CASE STUDY

Figure 1 shows two versions of the first 4 bars of Speed the Plough, a tune well-known across the British Isles (at the time of writing the abcnotation.com tune search has 244 tunes with a title which includes the phrase "Speed the Plough"). The first version in Fig. 1 is drawn from an English collection and the second, with the title "God Speed the Plough", from an Irish collection. Clearly these tunes are related but with distinct differences, particularly in the second and fourth bars.

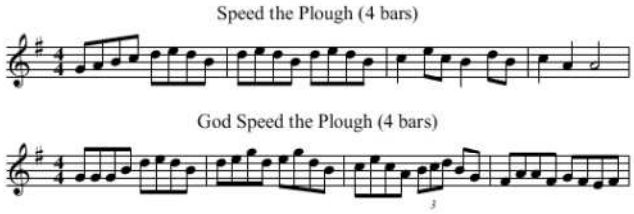


Figure 1. Two tune variants for Speed the Plough.

Subjectively, as a working musician, it is typical in tunes like this (which can be played as a reel, with an even rhythm, or a hornpipe, with a dotted rhythm), that the emphasis is placed on the odd numbered notes, and in particular the first note of each beam. The strongest notes of the bar are thus 1 and 5, followed by 3 and 7.

To capture this emphasis when matching tune variants it might be possible to use some sort of similarity metric which weights stress (so that matching 1st notes carry more importance than, say, 2nd notes, e.g. Typke, 2007). However, in this paper the approach is to build a multilevel (hierarchical) representation of the tunes.

The multilevel paradigm is simple one which involves recursive coarsening to create a hierarchy of increasingly coarse approximations some original representation. As a general solution strategy, the multilevel paradigm is widely used in combinatorial and other optimisation problems and can be extremely effective, both at imparting a global perspective and at accelerating solution techniques, (Walshaw, 2008).

Figures 2 and 3 show multilevel coarsened versions of the original tunes, where the weakest notes are recursively replaced by removing them and extending the length of the previous note by doubling it.

At level 0, i.e. the original, the tunes are quantised to show every note as a sixteenth note, thus simplifying the coarsening process. In addition the triplet in bar 3 of "God Speed the Plough" is simplified by representing it as two eighth notes, the first and last notes of the triplet.

To generate level 1, the 2nd, 4th, 6th and 8th notes are removed from each bar; for level 2, the original 3rd and 7th notes (which are now the 2nd and 4th) are removed; for level 3, the original 5th note (now the 2nd) is removed. As can be seen, as the coarsening progresses the two versions become increasingly similar and thus provide a good scope for melodic comparisons which ignore the finer details of the tunes.

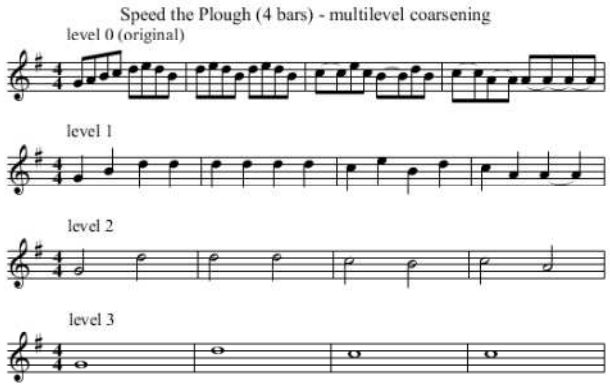


Figure 2. Multilevel coarsening of Speed the Plough

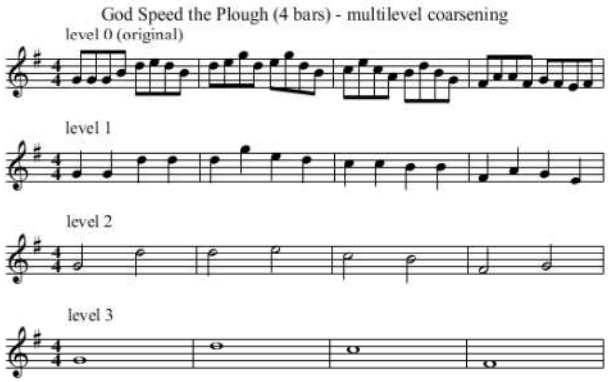


Figure 3. Multilevel coarsening of God Speed the Plough

The technique is related to, although somewhat more general than, the hierarchical contour pattern analysis technique proposed by Anagnostopoulou *et al*., (2013).

3. IMPLEMENTATION

The basis of the multilevel implementation is straightforward. Each tune is initially normalised and quantised and then recursively coarsened down to a skeleton representation with just one note per bar. Melodic matching can then take place at every level with heuristics used to optimise and enhance performance.

3.1 Normalisation

As part of the normalisation process, each tune is cleaned of grace notes, chords and other ornaments. Generally most tunes from the abc corpus are single-voiced (Walshaw, 2014), but if not, only the first voice is used for the matching.

Next, each tune is quantised so that longer notes are replaced with repeated notes (e.g. a half note is replaced with 4 eighth notes). In Western European folk music most tunes are written with eighth notes as the shortest note length; exceptions are, for example, polkas (in 2/4) and 3-time bourrées (3/8) which use sixteenth notes.

The shorter the note used for quantisation, the deeper the multilevel hierarchy and hence the more computationally complex the matching process. On the other hand quantising with longer notes can lose information, so the algorithm uses a long established heuristic in abc notation that if the time signature evaluates as fraction to 0.75 or above, tunes are quantised with eighth notes, otherwise sixteenth notes are used (for example, a tune in 3/4, which evaluates to 0.75, is quantised with eighth notes whereas a tune in 2/4 = 0.5 uses sixteenth notes).

If triplets (i.e. 3 notes in the space of 2) are encountered they are replaced by the first and third notes, but any other tuplets result in the tune being ignored (see Limitations, Section 3.4).

Figures 2 & 3 show examples of the quantisation, here labeled “level 0 (original)” which refers to the fact that no coarsening has yet taken place. In both examples, the tunes are rendered entirely as eighth notes and notice that the triplet `Bcd` in the third bar of God Speed the Plough has been replaced with the notes `Bd`.

3.2 Coarsening

The coarsening works by recursively removing “weaker” notes from each tune to give increasingly sparse representations of the melody. Coarsening progresses until there is one note remaining in each bar; it would be possible to go even further, coarsening down to one single note for a tune, but experimentation suggests that the bar is a good place to stop.

The choice of which notes to remove is subjective, based on the author’s experience of playing Western European folk music and in the current implementation the default coarsening strategy removes every other note, i.e. the off-beats (see Figures 2 & 3).

Exceptions to the “remove every other note” rule are handled with heuristics, typically for tunes in compound time. Thus for jigs in 6/8, 9/8 & 12/8, which are normally written in triplets of eighth notes, the weakest notes are generally the second of each triplet. As an example, for a slip jig in 9/8, once the tune is quantised to 9 eighth notes, the initial coarsening would remove notes 2, 5 & 8. A similar reasoning applies for waltzes, mazurkas and polskas in 3/4, so that for 3 quarter notes in a bar, the weakest is generally the second. The heuristics for dealing with these, and other less common time signatures (e.g. 5/4, 5/8, 7/8, 11/8, …), are summarised in Table 1 (although see also Section 3.4, Limitations).

Note that these heuristics are applied only once during the coarsening, specifically when the number of notes remaining in each bar is equivalent to the number of beats as expressed by the upper number of the time signature. For example, a tune in 3/8 would be quantised as 6 sixteenth notes which initially be coarsened by removing the notes 2, 4 & 6. This leaves 3 notes, 1, 3 & 5, equivalent to the number of beats and so, as indicated by the table, the second of these is removed.

Time signature	Beats per bar	Typical emphasis	Notes to remove
3/2, 3/4, 3/8	3	**1** 2 3	2
5/4, 5/8	5	**1**,2,3 **4**,5	2
6/4, 6/8	6	**1**,2,3 **4**,5,6	2 5
7/8	7	**1**,2,3 **4**,5 **6**,7	2
9/8	9	**1**,2,3 **4**,5,6 **7**,8,9	2 5 8
11/8	11	**1**,2 **3**,4 **5**,6,7 **8**,9 **10**,11	6
12/8	12	**1**,2,3 **4**,5,6 **7**,8,9 **10**,11,12	2 5 8 11

Table 1. Coarsening heuristics.

If the number of notes is not equivalent to the number of beats then evenly numbered notes are removed. For example with a slip jig in 9/8 with notes 1,2,3 4,5,6 7,8,9 at level 0, the initial coarsening using the heuristic from Table 1, removes 2, 5 & 8 leaving 1,3 4,6 7,9 at level 1. The next coarsening removes every other note, 3, 6 & 9, leaving 1, 4, 7 at level 2. This continues to give 1, 7 at level 3 and finally 1 at level 4.

Note that, in contrast to the motivational example coarsenings shown in Figures 2 & 3, this results in an unmusical result: each bar initially contains 9 eighth notes and 3 are removed so unless the remaining six notes are written as tuplets (specifically 2 notes in the time of 3), the tune no longer makes sense. However, since all notes are the same length (because of quantisation), this need not concern the implementation which is really only concerned with pitch rather than duration.

3.3 Similarity Measure

Once the multilevel representations are constructed a variety of methods could be used to actually compare tunes at each level (e.g. Kelly, 2012; Stober, 2011; Typke *et al.*, 2005). In the current implementation, each level is converted to intervals and then matching is done using the Longest Common SubString (LCSS) algorithm. However, in principle various methods can be used (this is a strength of the multilevel paradigm which is not generally tied to a particular local search strategy).

3.3.1 Multilevel Similarity

Because folk and traditional tunes can differ widely at the finest level whilst resembling each other at the coarser levels, a possibility for quantifying the similarity, S_{XY}, between a pair of tunes X and Y is simply to add up the lengths of all the LCSS values at every level. In other words, if S^l_{XY} expresses the similarity between tunes X and Y at level l then $S_{XY} = \Sigma_l S^l_{XY}$.

To illustrate this, Tables 2 & 3 show the semitone intervals for the two tunes in Figures 2 & 3 at all 4 levels. Notice that, because the coarsening is removing alternate notes at every level, every interval at levels 1, 2 & 3 is the sum of the two intervals in the parent level.

Applying the LCSS algorithm to all four levels (in reverse order) gives the following results:

- At level 3, the LCSS is 7,-2 so $S^3_{XY} = 2$
- At level 2, the LCSS is 7,0 so $S^2_{XY} = 2$
- At level 1, the LCSS is -5,-3,-2,-3 so $S^1_{XY} = 4$
- At level 0, the LCSS is 3,2,-2,-3,3,2 so $S^0_{XY} = 6$

Hence the similarity of this particular pair of tunes is quanitified as $S_{XY} = 2 + 2 + 4 + 6 = 14$. This puts value on the structural correspondences at levels 2 and 3 as well as the detailed similarities at levels 0 and 1.

Note that at levels 1 and 0 the LCSS are found at different positions in each tune. At level 1 it compares the notes `eBdcA` in Speed the Plough (bars 3 & 4) with the notes `BF#AGE` in God Speed the Plough (essentially the whole of bar 4), so is somewhat of a false positive as these two phrases are not really related (although see Section 3.3.3 for a way to avoid this). At level 0 it compares `BdedBde` from both tunes and is therefore more representative as the beamed notes `dedB` occur in both tunes (repeatedly in Speed the Plough) and are a characteristic feature of several other versions of the tune.

3.3.2 Multilevel Distance Metric

For convenience, it can be helpful to formulate the matching problem as a minimisation and hence to express the similarity as a distance, $D_{X,Y}$.

	Bar 1	Bar 2	Bar 3	Bar 4
Level 0	2 2 1 2 2 -2 -3 3	2 -2 -3 3 2 -2 -3 1	0 4 -4 -1 0 3 -3 1	0 -3 0 0 0 0 0
Level 1	4 3 0 0	0 0 0 -2	4 -5 3 -2	-3 0 0
Level 2	7 0	0 -2	-1 1	-3
Level 3	7	-2	0	

Table 2. Multilevel interval analysis for Speed the Plough.

	Bar 1	Bar 2	Bar 3	Bar 4
Level 0	0 0 4 3 2 -2 -3 3	2 3 -5 2 3 -5 -3 1	4 -4 -3 2 3 -3 -4 -1	3 0 -3 1 -1 -2 2
Level 1	0 7 0 0	5 -3 -2 -2	0 -1 0 -5	3 -2 -3
Level 2	7 0	2 -4	-1 -5	1
Level 3	7	-2	-6	

Table 3. Multilevel interval analysis for God Speed the Plough.

This is easy to do by computing, at each level l,

$$D^l_{XY} = \min(\text{length}(X^l), \text{length}(Y^l)) - S^l_{XY},$$

where length(X^l) is the length of the array of intervals at level l. Then $D_{XY} = \Sigma_l D^l_{XY}$.

In the case of the two example tunes, length(X^l) = length(Y^l) at every level and so the partial distances are given by:

- At level 3, $D^3_{XY} = 3 - S^l_{X,Y} = 1$
- At level 2, $D^2_{XY} = 7 - S^2_{XY} = 5$
- At level 1, $D^l_{XY} = 15 - S^l_{XY} = 11$
- At level 0, $D^0_{XY} = 31 - S^0_{XY} = 25$

and so the total distance is $D_{XY} = 1 + 5 + 11 + 25 = 42$.

3.3.3 Algorithmic and Other Variants

Although essentially expressing the inverse of S_{XY}, D_{XY} puts much more emphasis on the finer levels, simply because the arrays are much longer. This means that a pair of tunes which match closely at the coarsest levels may still have a relatively large D_{XY} value if the LCSS at the finest level is relatively short.

To compensate for this, a possibility is to normalise the coarser levels so that the length of the LCSS at every level has approximately the same contribution (as it does for S_{XY}). Since, in most cases, the length of the interval arrays is halved at each successive level, the simplest way to do this is just to multiply D^l_{XY} by 2^l so that

$$\underline{D}^l_{XY} = 2^l \,.\, [\min(\text{length}(X^l), \text{length}(Y^l)) - S^l_{XY}]$$

For the two example tunes this gives:

- At level 3, $\underline{D}^3_{XY} = 8 \,.\, [3 - S^l_{X,Y}] = 8$
- At level 2, $\underline{D}^2_{XY} = 4 \,.\, [7 - S^2_{XY}] = 20$
- At level 1, $\underline{D}^l_{XY} = 2 \,.\, [15 - S^l_{XY}] = 22$
- At level 0, $\underline{D}^0_{XY} = 1 \,.\, [31 - S^0_{XY}] = 25$

and so the total distance is $\underline{D}_{XY} = 8 + 20 + 22 + 25 = 75$. Although this is larger in absolute value than D_{XY}, because of the weighting, it can often better distinguish between matches.

To illustrate this further it is helpful to introduce another tune, a Northumbrian version of Speed the Plough, as shown in Figure 4, this time with an 8 bar fragment of the tune in a different key and with an anacrusis. This is a much closer match to the original "Speed the Plough" than "God Speed the Plough" is.

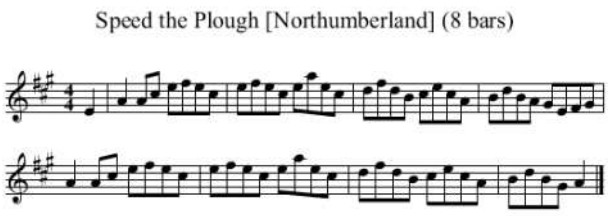


Figure 4. A further tune variant for Speed the Plough.

Denoting "Speed the Plough [Northumberland]" as tune Z (where X refers to "Speed the Plough" and Y to "God Speed the Plough"), the similarity values are $S_{XZ} = 21$ and $S_{YZ} = 18$ as compared with $S_{XY} = 14$. The distance values are $D_{XZ} = 35$ and $D_{YZ} = 38$ as compared with $D_{XY} = 42$ and the normalised distance values are $\underline{D}_{XZ} = 57$ and $\underline{D}_{YZ} = 71$ as compared with $\underline{D}_{XY} = 75$ showing that the normalised distance, $\underline{D}$, gives a better spread than just the distance, D, and clearer discrimination.

Another possible algorithmic variant is the way in which tunes of different lengths are compared. Using the minimum of the two lengths, i.e. min(length(X^l), length(Y^l)), instead of the maximum to calculate the distance D^l_{XY} from the similarity S^l_{XY} seems sensible, particularly in a search context where the query is likely to be a tune fragment. However, in other settings it might not be appropriate and this is explored experimentally in a sister paper (Walshaw, 2015).

With regard to the matching at each level, a potential pitfall of the LCSS algorithm is that it may give false positives by matching short phrases from completely different parts of the tune. One way around this is to include bar markers or even bar numbering within the strings that are to be matched. So, for example, level 2 of "Speed the Plough" which would normally be represented as "7,0,0,-2,-1,1,-3" (see Table 2) could instead be represented as "7,0,|,0,-2,|,-1,1,|,-3" where the "|" symbols represent bar lines. This means that any matched common substrings must respect bar lines (unless they are shorter than the length of a bar).

Furthermore, if the bar symbols are numbered, e.g. "7,0,$|^1$,0,-2,$|^2$,-1,1,$|^3$,-3", then matched common substrings must also respect the position in the tune. (If matching of subsections of the tune is important then the numbering can be restarted at natural breaks such as double bar lines and repeat marks; however, that is not been tested here.)

In terms of implementation, the "strings" of intervals are represented as an array of short integers so that bar markers (or numbers) can easily be included with large integer values outside the possible range of intervals.

This inclusion of bar markers or numbers is more of a representational variant than an algorithmic one and does increase the computational complexity of the multilevel matching (as the strings to be compared by the LCSS algorithm are longer). However, it has a significant effect on the results and is an important component of the multilevel LCSS algorithm.

Finally, the multilevel framework also allows for the use of optimisation heuristics to terminate the matching process early, at the coarser levels, when it looks unpromising. This is discussed in more detail alongside the experimentation (see Section 4.5).

3.4 Limitations

The current implementation is a prototype and there are a number of tune features it cannot handle fully. Some are treated as exceptions, in which case the tune is excluded from the search data, and in most cases it should be possible to improve the handling in future versions:

- **Tunes with no time signature** are treated as exceptions and excluded for the time being. An easy option is to include them and just coarsen by removing alternate notes for, say, 4 or 5 levels.
- **Tunes which change time signature** are treated as exceptions and excluded for the time being. This is

just an implementation issue and in principle the methods should work by taking account of changes and applying the coarsening methods accordingly.

- **Complex time signatures / emphasis patterns** may mean that the heuristics in Table 1 are incorrect. For example, some tunes in 5/4 have a **1**,2 **3**,4,5 emphasis pattern and there are Eastern European tunes in 9/8 with **1**,2,3 **4**,5 **6**,7 **8**,9. It is probably impossible to identify all of these anomalies accurately and anyway these are not common in the corpus (which is mostly comprised of Western European and North American tunes) so they treated as any other tune.
- **Tuplets –** other than triplets, tuplets are not yet handled correctly and any tune with tuplets is excluded.
- **Short bars** (e.g. anacruses & repeat bars) are treated like any other bar meaning that the Table 1 heuristics may not be applied correctly if the bar is shorter than it would normally be.
- **Polyphonic tunes** are currently handled by extracting the first voice. In principle it should be possible to extract each voice and treat it as a separate tune, but this is not yet implemented.
- **Repeat signs** are currently ignored so that two identical tunes, one with repeat signs and one written out in full would not have a distance measure of 0. In principle it would be possible to expand all repeats, but this would significantly add to the computational complexity of the matching.
- **Transcription errors** and extraneous symbols in the abc code are mostly ignored but if the parser really cannot understand the input, the tune is excluded.

4. EXPERIMENTATION

4.1 Experimental Framework

The experimental framework uses two illustrative search queries to assess and demonstrate some of the aspects of the multilevel search algorithm. Both are chosen as examples of tunes which are very well known in the Western European folk canon and which therefore have many variants represented in the abc music corpus.

The first search query is for the first two bars of "Speed the Plough" (see Figure 1), a tunethought to have been composed by John Moorehead, originally from Edinburgh, in around 1800 and named after its inclusion in an eponymous play[1].

The second search query is the first two bars of the tune "Black Joker" (also known as "Black Joke", "Black Jack", "Black Jock", "Black Joak", etc., as well as "But the House and Ben the House" in the Shetland Isles, "Sprig of Shillelagh" in Ireland and "La Badine" in the Netherlands). This is an older tune, dating from at least the early eighteenth century, and still popular for morris dancing[2]. Figure 5 shows the first 4 bars of a number of variants (all drawn from the test dataset, Section 4.3), as well as the search query.

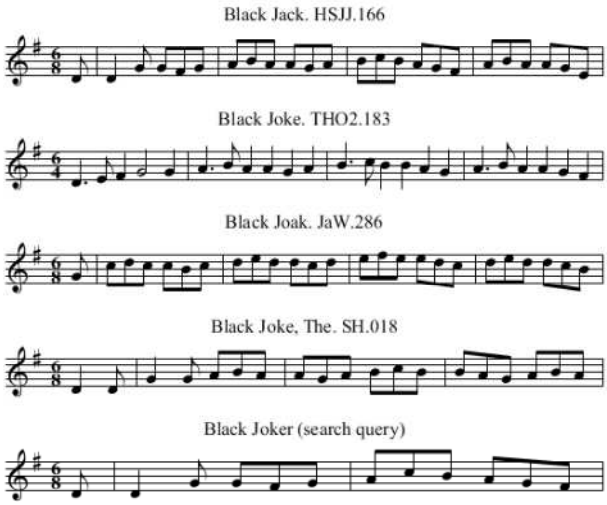


Figure 5. Several variants of the incipit for Black Joker.

The queries are run on two datasets: a small set of test data (5,610 tunes) and the entire corpus (460,000 tunes, reduced to 168,960 after various groups are excluded).

4.2 Variants Tested

A number of algorithmic and representational variants are tested. In particular:

- **Normalisation**: as described in Section 3.3.3 this weights the distance measure so that matching substrings of equal length at different levels of the multilevel representation contribute approximately the same amount to the distance measure.
- **Bar markers/numbers:** as described in Section 3.3.3 bar markers or numbers can be included in the "strings" of intervals matched at each level. This forces the LCSS algorithm to respect the position of notes relative to bar lines (if using markers) or additionally the position of notes within the tune (if using numbers).
- **Single-level search:** in order to provide a comparison, all the tests are also performed with single level searches, i.e. so that the LCSS algorithm is just used at level 0, ignoring the multilevel representation.

4.3 Results – Test Dataset

The initial experimentation uses a small subset of the full corpus consisting of the 5,638 abc transcriptions taken from www.village-music-project.org.uk[3], a collection of social dance music from England mostly transcribed from hand-written manuscript books in museums and library archives. Of these 28 are removed due to implementation limitations (see Section 3.4) leaving 5,610.

One of the advantages of using this dataset is the diversity of the material, sometimes with several versions for each tune. Another advantage is that the tunes include the manuscript they are taken from abbreviated in the tune title, making it easy to identify specific instances.

[1] See http://www.ibiblio.org/fiddlers/speed.htm for a comprehensive history

[2] See http://www.ibiblio.org/fiddlers/BLACK.htm

[3] See http://village-music-project.org.uk/

4.3.1 Search Metrics

Testing a search algorithm, particularly in a corpus where there is no "ground truth" established, is not an easy matter. In the testing below, the aim is to find all the versions of the tune with matching titles, solely by using the musical representation. In other words, the testing considers tunes which contain "Speed the Plough" or "Black Jack/Joke/Joke" (because of the variant spellings) in the title and where they appear in the search results.

In the case of the test data there are 10 instances with "Speed the Plough" in the title and 9 instances with "Black Jack" / "Black Joak" and "Black Joke" (plus 2 with "Sprig of Shillelah" and "Sprig of Shillela" which are not considered). These two sets of instances are each respectively referred to as **target sets** below.

One way of assessing the results is consider the entire dataset and how "distant" each target set is from the query. Two metrics are used for this purpose:

- **Avg. distance**: the simplest measure is just to average all the distances, D_{ST}, of tunes, T, in the target set from the search query, S, so that if {T} represents the target set, calculate $\Sigma_{\{T\}} D_{ST} / |\{T\}|$. However, since the various algorithmic variants can have different scales for the distance calculation, the distances are all normalised by dividing by the maximum possible distance to give values between 0 and 1.
- **Halfway index**: another way to assess the results is to consider how many tunes would need to be shown in the search results before the entire target set appeared. However, this measure could be badly influenced by outliers in the target set (which have a matching tune title but which are not similar musically) so a more robust metric is to consider how many tunes need to be shown in the results before half of the target set appears. This is known as the **halfway index**.

Another way of assessing the results is to use the distance measure to identify a **results set** of tunes close to the search query, i.e. where $D_{ST} < \Delta$ for some chosen value Δ. The results set then contains the tunes offered to the user performing the search. Two further metrics are used to assess the results set:

- **Size**: ideally the results set should be small enough to be useable, but large enough to contain some diversity (e.g. similar tunes that the user might not be aware of).
- **Instances**: a count of how many instances from the target set are contained in the results set.

4.3.2 Results

Table 4 shows the results for the two search queries in the test dataset. The first 3 columns indicate the algorithmic variant (as described in Section 4.2), and in particular: the search type, either multilevel (ML) or single-level (SL); whether normalisation is applied to multilevel distances give each level equal weight; whether bar markers or numbers are included in the tune representation. The results are in the last 4 columns with the "best" values indicated in boldface.

Variant			Target set		Results set (Δ = 0.5)	
Type	Normalisation	Bars	Halfway index	Avg. distance	Size	Instances
Speed the Plough						
ML	yes	numbered	**5**	**0.187**	264/5,610	**10/10**
ML	yes	marked	6	**0.187**	2,612/5,610	**10/10**
ML	yes	ignored	9	0.265	1,310/5,610	**10/10**
ML	no	numbered	6	0.382	16/5,610	8/10
SL	-	numbered	94	0.659	4/5,610	3/10
Black Joker						
ML	yes	numbered	**12**	0.364	290/5,610	6/9
ML	yes	marked	19	**0.299**	2,829/5,610	**7/9**
ML	yes	ignored	24	0.378	1,387/5,610	**7/9**
ML	no	numbered	34	0.500	19/5,610	4/9
SL	-	numbered	187	0.650	6/5,610	3/9

Table 4. Results for the test dataset.

As can be seen, the tables suggest that the multilevel search with normalisation (top 3 rows of results) generally offers the best results in terms of the halfway index, the average distance and the number of target instances in the results set.

The single level scheme means essentially just using LCSS as a search strategy and substantially increases the average distance. This could just be an issue of scaling in the distance function but it also hugely increases the halfway index, fundamentally because single-level LCSS is an exact search with none of the approximate matches provided by the coarser levels in the multilevel versions. As a consequence the results set is tiny and does not contain many target instances (although it might be possible to tailor the value of Δ to improve this).

The multilevel results without normalisation seem to confirm these findings. Normalisation weights the distance contribution to give approximately equal emphasis from each level and if it is not used the distance measure is heavily biased towards the matching at the finest (original level), so it is not too dissimilar from using a single-level LCSS search. Nonetheless, the contributions from coarser levels do seem to discriminate further between different tunes and hence improve the results.

The choice of whether to use bar markers / bar numbering could easily be left to the user and, in particular, bar numbering only makes sense if the search query is known to come from the start of the tune (although it could be adapted to apply to the start of each section of the tune or for user-chosen numbering). However, the results make it clear that it is significantly better to use bar marking or numbering than to ignore the bar lines.

Overall the best results seem to come from using the multilevel algorithm with normalisation and bar numbering. This configuration produces half of the target in-

stances in the first few results (as measured by the halfway index) and the results set for $\Delta = 0.5$ contains a few hundred tunes (around 5% of the dataset). The results set does not always contain all of the target instances but that is because not all of the targets are particularly close matches – in particular the 3 missing for the Black Joker query are shown in Figure 5 with the manuscript identifiers THO2.183, JaW.286 & SH.018; of these the first has, unusually, been transcribed in 6/4, the second is not a close match over the first 2 bars (although would have been if the search query were 4 bars long), whilst the third has been transcribed (on the original manuscript) with the bar lines in the wrong place.

If bar marking is used instead of numbering, the size of the results set increases by a factor of around 10 in both cases, mostly likely as the result of a large number of false positive matches from elsewhere in the tune. These enlarged results sets also account for the fact that they contain the highest number of target set instances.

To tease this out a little further it is worthwhile studying the histograms of distances, shown in Figure 6, here for Speed the Plough, although those for Black Joker are similar (note that the scaling on the *x*-axis is different when bars are ignored as the representation strings are shorter and hence the maximum distance is smaller).

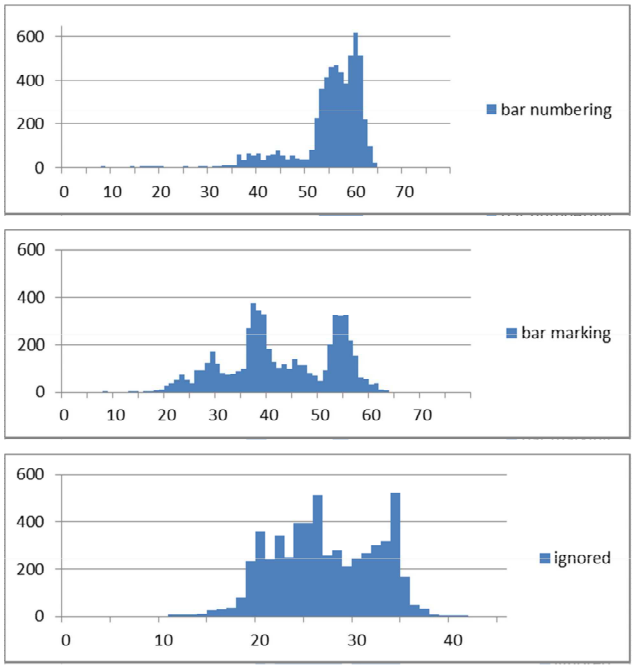


Figure 6. Distance histograms for Speed the Plough.

These histograms show the extent of the results set: for bar marking/numbering, any tune within a distance of 39 ($\Delta = 0.5$) is included in the results set. When bars are ignored the distance is 22.

What is clear from them is that bar numbering is better at separating out the results with a gradual accumulation of matching tunes up to the value 50 (corresponding to $\Delta \approx 0.625$) and the main bulk of tunes coming after that. In contrast, bar marking or ignoring bar lines altogether, spreads the distance results out more evenly and so the results set can grow rapidly as Δ is increased.

4.4 Results – Entire Corpus

The second data set is the entire abc corpus which at the time of writing consists of around 460,000 tunes from across the web (Walshaw, 2014). Of these 244,000 are exact electronic duplicates which are excluded and another 40,000 are potentially copyright and also ignored. A further 7,000 are excluded because of implementation limitations (see Section 3.4), leaving a total of 168,960 used for testing.

The target sets in this case (i.e. those tunes which have closely matching titles) contain 84 versions of Speed the Plough and 88 versions of Black Joker.

Table 5 shows the results for the full corpus set out as previously. These seem to bear out the findings established for the test dataset – the single-level and multilevel with no normalisation give the worst results and, in the multilevel results with normalisation, using bar numbering provides the best results in terms of the halfway index and nearly the best results in terms of the number of target instances in the results set (once again, the best results arise from results sets which are many times bigger).

Variant			Target set		Results set ($\Delta = 0.5$)	
Type	Normalisation	Bars	Halfway index	Avg. distance	Size	Instances
Speed the Plough						
ML	yes	numbered	**91**	0.242	6,651/168,960	69/84
ML	yes	marked	96	**0.216**	71,163/168,960	**73/84**
ML	yes	ignored	117	0.271	37,171/168,960	69/84
ML	no	numbered	129	0.360	318/168,960	60/84
SL	-	numbered	341	0.527	168/168,960	41/84
Black Joker						
ML	yes	numbered	**159**	0.326	7,688/168,960	58/88
ML	yes	marked	273	**0.286**	77,734/168,960	**64/88**
ML	yes	ignored	314	0.332	36,209/168,960	**64/88**
ML	no	numbered	201	0.434	397/168,960	53/88
SL	-	numbered	2882	0.582	179/168,960	35/88

Table 5. Results for the full corpus.

4.5 Optimisations

LCSS is computationally expensive, especially since, in the multilevel context, it may be carried out at 4 or sometimes 5 levels. However the multilevel framework also allows for the use of optimisation heuristics to terminate the matching early, at the coarser levels, when it looks unpromising and this is used to significantly speed up the matching process, as follows.

4.5.1 Distance Threshold Optimisation

If a distance threshold, Δ, is set it provides a natural early termination condition when comparing tunes, so that as each D^{l}_{XY} contribution is added in to the distance meas-

ure D_{XY}, the matching can be ended as soon as the partial sum $\Sigma_l D^l_{XY}$ is greater than the current distance threshold. Since LCSS is an $O(n^2)$ operation this can result in significant savings in computational complexity, especially if, as is typical in the multilevel framework (Walshaw, 2008), the matching is carried out coarsest to finest with the shortest arrays compared first.

4.5.2 Coarse Level Matching Limit (CLML)

Distance threshold optimisation will not change the results set, but if the desire is refine the search accuracy an option is to exclude any tune from the results if the length of the LCSS for the coarsest level L is less than some minimum matching limit M^L or, in other words, $S^L_{XY} < M^L$. With a short search query of 2 bars, as used here, it makes sense to set $M^L = 2$.

Since the coarsest level has one note per bar and the final note remaining in each bar during coarsening is always the first note in the bar, this is effectively the same as saying that candidate tunes must have the same first note as the search query in each bar, for its entire length.

Note also that, in principle, it would be possible to employ this heuristic at any level by setting minimum matching limits for finer levels, e.g. M^{L-1}, M^{L-2}, etc. However this has not been tested.

4.5.3 Results

Table 6 shows the results of the two optimisations applied the multilevel algorithm (using normalisation and bar numbering). Here the runtime gives the time, in milliseconds to search through 168,960 multilevel tune representation (although not to read the tunes in from file).

	Target set		Results set		
Optimisation	Halfway index	Avg. distance	Size	Instances	Runtime (ms)
Speed the Plough					
none	91	0.242	6,651	69/84	1,295
threshold	91	0.242	6,651	69/84	822
CLML	**45**	n/a	5,912	69/84	**250**
Black Joker					
none	159	0.326	7,688	58/88	1,423
threshold	159	0.326	7,688	58/88	853
CLML	**70**	n/a	7,330	57/88	**400**

Table 6. Optimisation results.

As can be seen, the distance threshold optimisation does not change the results qualitatively, but does reduce the runtime by over a third. In contrast the CLML optimisation, by excluding tunes which do not completely match the search query at the coarsest level, not only reduces the runtime still further, but also significantly improves the halfway index of the target sets.

5. CONCLUSION

This paper has presented a multilevel algorithm for matching tunes when performing inexact searches in symbolic musical data.

The basis of the algorithm is very straightforward: each tune is initially normalised and quantised and then recursively coarsened, typically by removing weaker offbeats, until the tune is reduced to a skeleton representation with just one note per bar. Melodic matching can then take place at every level, coarse to fine.

The matching implemented here uses the Longest Common SubString (LCSS) algorithm, but in principle a variety of similarity measures could be used.

The multilevel framework allows inexact matches to occur by identifying similarities at course levels and is also exploited with the use of early termination heuristics at coarser levels, both to reduce the computational complexity and enhance the matching qualitatively.

Initial results suggest that the algorithm, coupled with the LCSS, works very well at performing searches on a corpus of tunes. However, this is only a prototype and further work remains to be done, in particular:

- further testing with a broader, more diverse range of search queries;
- exploration of different similarity measures within the algorithm;
- better handling of exceptions and anomalies (see Section 3.4) to improve the robustness.

6. REFERENCES

Anagnostopoulou, C., Giraud, M., & Poulakis, N. 2013. Melodic contour representations in the analysis of children's songs. In P. van Kranenburg *et al.*, Eds., *3rd Intl Workshop on Folk Music Analysis*pp. 40–43. Amsterdam: Meertens Institute and Utrecht University.

Kelly, M. B. 2012. *Evaluation of Melody Similarity Measures*. Queen's University, Kingston, Ontario.

Stober, S. 2011. Adaptive Distance Measures for Exploration and Structuring of Music Collections. In *Audio Engineering Society Conf.*, pp. 1–10.

Typke, R. 2007. *Music Retrieval based on Melodic Similarity*.

Typke, R., Wiering, F., & Veltkamp, R. C. 2005. A survey of music information retrieval systems. In *Proc. ISMIR*, pp. 153–160.

Walshaw, C. 2008. Multilevel Refinement for Combinatorial Optimisation: Boosting Metaheuristic Performance. In C. Blum, Ed., *Hybrid Metaheuristics - An emergent approach for optimization*pp. 261–289. Springer, Berlin.

Walshaw, C. 2014. A Statisical Analysis of the Abc Music Notation Corpus. In A. Holzapfel, Ed., *4th Intl Workshop on Folk Music* Analysis, pp. 2–9. Istanbul: Bogaziçi University.

Walshaw, C. 2015. A Multilevel Melodic Similiarity Framework. In *Proc. ISMIR (submitted).*

CHARACTERIZING COMPLEX NOTE TEMPORAL PROFILES WITH THE PLCA-HMM METHOD

Yuancheng Wang, Cazau Dorian and Olivier Adam
Sorbonne Universités / CNRS, UPMC Univ. Paris VI, UMR 7190
Institut Jean Le Rond d'Alembert, Equipe Lutherie-Acoustique-Musique, F-75005, Paris, France
{cazau,adam}@lam.jussieu.fr

1. EXTENDED ABSTRACT

To improve transcription performance, Automatic Music Transcription (AMT) systems integrate musical knowledge specific to the instrument repertoires under study, either during their parameter estimations or in a post-processing stage. The Probabilistic Latent Component Analysis (PLCA) is a powerful probabilistic framework to model audio signals and retrieve their main pitch components. It has been applied for AMT and source separation. However, this model still has essential problems for what concerns the task of AMT. In this project, we deal with the problem of extracting note duration. One of other drawbacks of current transcription systems is that in most cases, the non-stationarity of music sounds is not addressed. However, a note produced by a musical instrument can be expressed as a sequence of sound states, for example attack, transient, sustain, and decay parts (Bello et al., 2005).

In this project, we first analyse the PLCA problems in a statistical view, then we review the Hidden Markov Models PLCA model proposed by past studies (Mysore, 2010; Benetos & Dixon, 2013). This model is able to characterize note temporal profile through several temporal states, constraining state evolution with a left-to-right HMM topology. Compared with the baseline PLCA, this extended model recognizes much better the note onset. We apply this model to complex note temporal profiles, in particular those presenting strong amplitude modulation, and put into evidence the limits of this model in retrieving the proper duration of such notes. We propose in this project to replace the template learning part of this method by an optimized k-means, based on a similarity measure of the successive spectra in the note, and further improve the AMT performance through a more complex unsupervised time-frequency structure extraction.

Keywords : Automatic Music Transcription, HMM-PLCA, K-means, Non-eurogenetic music, Statistical learning, Computational Ethnomusicology

2. REFERENCES

Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., & Sandler, M. B. (2005). A tutorial on onset detection in music signals. *IEEE Trans. on Speech and Audio Proc.*, *13*, 1035–1047.

Benetos, E. & Dixon, S. (2013). Multiple-instrument polyphonic music transcription using a temporally constrained shift-invariant model. *J. Acoust. Soc. Am.*, *133*, 1727–1741.

Mysore, G. J. (2010). *A non-negative framework for joint modeling of spectral structure and temporal dynamics in sound mixtures*. PhD thesis, Stanford University, USA.

VIBRATO CHARACTERISTICS AND FREQUENCY HISTOGRAM ENVELOPES IN BEIJING OPERA SINGING

Luwei Yang, Mi Tian and Elaine Chew
Centre for Digital Music, Queen Mary University of London
{l.yang, m.tian, elaine.chew}@qmul.ac.uk

1. INTRODUCTION

Beijing opera, also known as Peking opera, is the predominant opera genre in China. The compound art form comprises of singing, instrumental playing, and acting. There is now growing interest in the study of Beijing opera from a computational perspective in recent years. Repetto & Serra (2014) created a dataset of sung Beijing opera melodies for computational analysis. Tian et al. (2014) investigated onset detection for Beijing opera percussion instruments using non-negative matrix factorization. Srinivasamurthy et al. (2014) utilized a Hidden Markov Model to transcribe and recognize percussion patterns. Sundberg et al. (2012) studied acoustically the singing for two Beijing opera roles, Laosheng and Dahualian, and found that the singing sound pressure level and pitch are higher than that of speech; furthermore, the vibrato rate was reported to be around 3.5Hz, which is lower than that generally found in Western classical singing. However, research into the expressivity of the featured singing and instrumentation in Beijing opera is still lacking in the literature.

Oriental opera singing possesses characteristics distinct from Western opera. This study aims to investigate the expressive characteristics of Beijing opera singing focusing on pitch and vibrato. Like the Sundberg et al. study, we will also focus on two major Beijing opera roles, those of Laosheng and Zhengdan. Our study involves a larger dataset, 16 performances instead of 7. While Sundberg et al.'s study focused on sound pressure level, pitch, and long-term average spectra, with vibrato being a side discussion, we will consider frequency distributions and details of vibrato parameters such as rate, extent, and sinusoid similarity.

We choose to focus on pitch because it is one of the most important features in computational music analysis; this is also true for analysis of world music. For example, Koduri et al. (2012) examined pitch histograms in Indian Carnatic music. Similar research is lacking for Beijing opera singing. Moreover, Chinese traditional music relies on a system with unique characteristics different from that of other music cultures (Tian et al., 2013).

Vibrato is one of the most important ornamental performing techniques in Beijing opera (Wichmann, 1989). Investigation into the nature of vibrato use in Beijing opera singing will thus assist in the understanding of the overall plots and the motifs of the story and the roles. We are also interested in investigating the relationship between vibrato use in singing vs. in the main instrument, i.e. jinghu, in Beijing opera. Jinghu is the predominant instrument and frequently doubles the singing voice in Beijing opera, but we lack vibrato statistics for the instrument. The jinghu and erhu belong to the same two-stringed instrument family, and possess similar form and, we anticipate, vibrato characteristics. In order to compare Beijing opera singing and jinghu vibrato characteristics, we use data previously established for the erhu (Yang et al., 2013).

2. DATASET

We used the singing voice dataset [1] described in (Black et al., 2014). This study is focused on two Beijing opera roles: Laosheng (老生) and Zhengdan (正旦) (also known as Qingyi (青衣)).

The dataset consisted of a total of 16 monophonic performances of well-known phrases in the opera sung by 6 different Chinese opera singers. All vibratos in each performance were labelled by the first two authors using Tony [2]. We found a total of 344 vibratos from the Laosheng role, and 273 vibratos from the Zhengdan role.

3. RESULTS

Figure 1 shows the smoothed frequency histogram envelopes for the Laosheng and Zhengdan data, respectively. The sung frequencies were extracted using the Tony software. The extracted frequencies were summed into one-cent bins, and the results were smoothed to obtain the histogram envelope.

Because Zhengdan is a female role, we expect the part's pitches to be higher than those of the Laosheng role. The results show that this is indeed the case; however, Laosheng's phrases utilize a wider pitch range than that of Zhengdan, and the highest pitches are in fact higher than most of Zhengdan's pitches. This maybe helpful in role type classification. It is interesting to note that the peaks in the frequency plot show that the Chinese opera melodies also use a semitone scale like that in Western music, although the most prevalent pitches use the traditional Chinese pentatonic scale.

Figure 2 compares the distributions of the vibrato rates, extents, and sinusoid similarity measures for the recorded

[1] http://www.isophonics.net/SingingVoiceDataset
[2] https://code.soundsoftware.ac.uk/projects/tony/files

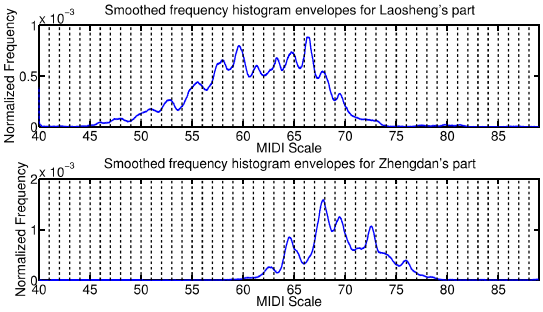


Figure 1: Smoothed frequency histogram envelopes for Laosheng's and Zhengdan's parts.

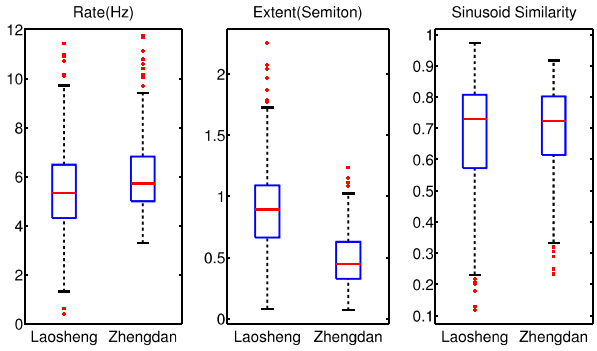


Figure 2: Box plots of vibrato statistics (rate, extent, sinusoid similarity) for Laosheng and Zhengdan.

singing for the Laosheng and Zhengdan roles. The vibrato parameters were obtained using methods described in (Yang et al., 2013). The plots show that the Laosheng and Zhengdan roles are sung with similar vibrato rates ranges, similar to vibrato employed in Western opera singing. This observation contradicts the findings of (Sundberg et al., 2012), in which the vibrato rate was shown to be about 3.5Hz. The singing vibrato rate range is also similar to that of erhu described in (Yang et al., 2013)

The Laosheng role was sung with vibrato having an average extent of almost one semitone, while Zhengdan's role is sung with mean vibrato extent approximately half a semitone wide. These extent values are consistent with those in Western opera singing, which ranges from 0.34 to 1.23 semitones (Prame, 1997). The vibrato extent range is similar to that found for the erhu in (Yang et al., 2013).

The vibrato sinusoid similarity for both roles have a median of 0.71, which is lower than the 0.85 found in erhu playing (Yang et al., 2013). This indicates that singers may have more expressive freedom than erhu players, leading to more non-sinusoid shaped vibratos. Also, it may be more difficult for a singer to control the voice in a perfectly periodic manner.

4. CONCLUSION

In this study, we examined the pitch histograms and vibrato statistics of singing of two Beijing opera roles, Laosheng and Zhengdan. Laosheng (a male role) employs lower pitches, but a much larger pitch range, than Zhengdan (a female role). The singing employed all twelve pitches in the Western scale, and the vibrato rates and extents are similar to those in Western opera singing. Laosheng's role is sung with significantly larger vibrato extents than that of Zhengdan.

Some interesting phenomena have been noticed. Vibratos in Beijing opera singing have rates and extents similar to that of erhu performance, which we use as a proxy for characteristics of jinghu vibratos. This is a demonstration, to some degree, of the hypothesis that string instruments are designed to mimic the human voice. The similarity between jinghu and erhu vibratos should be confirmed in future work.

Finally, portamenti also form an important characteristic of Beijing opera singing. Future work will include statistical analysis of portamenti properties in Beijing opera performance.

5. REFERENCES

Black, D. A. A., Ma, L., & Tian, M. (2014). Automatic identification of emotional cues in Chinese opera singing. In *Proc. of 13th Int. Conf. on Music Perception and Cognition and the 5th Conference for the Asian-Pacific Society for Cognitive Sciences of Music (ICMPC 13-APSC0M 5).*

Koduri, G. K., Serrà, J., & Serra, X. (2012). Characterization of intonation in carnatic music by parametrizing pitch histograms. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR).*

Prame, E. (1997). Vibrato extent and intonation in professional western lyric singing. *J. Acoust. Soc. Am.*, *102*, 616–621.

Repetto, R. C. & Serra, X. (2014). Creating a corpus of jingju (Beijing opera) music and possibilities for melodic analysis. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR).*

Srinivasamurthy, A., Repetto, R. C., Sundar, H., & Serra, X. (2014). Transcription and recognition of syllable based percussion patterns: The case of Beijing opera. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR).*

Sundberg, J., Gu, L., Huang, Q., & Huang, P. (2012). Acoustical study of classical Peking opera singing. *Journal of Voice*, *26*(2), 137–143.

Tian, M., Fazekas, G., Black, D., & Sandler, M. (2013). Towards the representation of Chinese traditional music: A state of the art review of music metadata standards. In *Proc. of the International Conference on Dublin Core and Metadata Applications.*

Tian, M., Srinivasamurthy, A., Sandler, M., & Serra, X. (2014). A study of instrument-wise onset detection in Beijing opera percussion ensembles. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing.*

Wichmann, E. (1989). *Listening to theatre: the aural dimension of Beijing opera.* University of Hawaii Press.

Yang, L., Chew, E., & Rajab, K. Z. (2013). Vibrato performance style: A case study comparing erhu and violin. In *Proc. of the 10th International Conference on Computer Music Multidisciplinary Research (CMMR).*

Authors

Credit photos on the cover page from Marc Chemillier, EHESS, 2015
Directed by Olivier ADAM and Dorian CAZAU
Designed by Olivier ADAM
Edited by the Association DIRAC, Paris, France
ISBN 979-10-95209-00-3