Fotografía Area de Turismo, Ayuntamiento de Málaga

#FolkMusic2017

Proceedings of the 7th International Workshop on Folk Music Analysis

ATIC
Grupo de Aplicación de las Tecnologías de la Información y Comunicaciones

Ingeniería de Comunicaciones

Escuela Técnica Superior de Ingeniería de Telecomunicación

UNIVERSIDAD DE MÁLAGA

ANDALUCÍA TECH
Campus de Excelencia Internacional

Vicerrectorado de Investigación y Transferencia

Málaga Convention Bureau

FMA 2017

Proceedings of the 7th International Workshop on Folk Music Analysis

14–16 June 2017, Málaga, Spain

EDITED BY: ISABEL BARBANCHO
LORENZO J. TARDÓN
ALBERTO PEINADO

ATIC
Grupo de Aplicación de las
Tecnologías de la Información
y Comunicaciones

Ingeniería de
Comunicaciones

Escuela Técnica Superior
de Ingeniería de
Telecomunicación

ANDALUCÍA TECH
Campus de Excelencia Internacional

UNIVERSITAS MALACITANA
UNIVERSIDAD
DE MÁLAGA
uma.es

Málaga
Convention
Bureau

ISBN: 978-84-697-2303-6

FMA 2017
7th International Workshop on
Folk Music Analysis
Málaga (Spain), 14-16 June 2017

Preface

Dear colleagues,

We are delighted to welcome you in Málaga, Spain, where we are going to celebrate the 7th International Workshop on Folk Music Analysis (FMA) from 14 to 16 June 2017.

This International Workshop brings together researchers from the field of ethnomusicology and the field of computational and analytical musicology. This Workshop will be a perfect framework to deal with topics related to musicology, ethnology, engineering and computer sciences including signal processing, pattern recognition, applied mathematics, etc.

FMA 2017 will provide the attendants with a fantastic forum to share research, thoughts, needs and discoveries between ethno-musicologists, musicians, librarians, students, museum curators, computer science experts and music information retrieval researchers to foster the creation of cross-disciplinary collaborative networks and the development of new interdisciplinary tools, methods, techniques and ideas to promote the enrichment of music, specially folk music, and the preservation and dissemination of World's musical cultural heritage.

This year FMA 2017 will feature 2 Keynote Talks, 1 Tutorial, 4 Oral Sessions and 2 Poster and Demo Sessions.

Besides this interesting Scientific Program, FMA 2017 also aims at giving the participants an unforgettable stay. In the evenings, several Social Activities will be provided. In all of them, the aim will be to promote interaction between participants and, at the same time, to show typical Folklore of Malaga, typical places and typical drinks and food.

We want to say thank you to all the people that have made FMA 2017 possible: the authors that sent their contributions, the reviewers, the FMA 2017 attendants, the FMA 2017 Conference Committee and, specially, the FMA 2017 Local Committee. We are very grateful to our sponsors Universidad de Málaga-Andalucia Tech, Departamento de Ingeniería de Comunicaciones, E.T.S.I. Telecomunicación and to our collaborators Málaga Convention Bureau and Ayuntamiento de Málaga.

We have worked hard preparing everything so that FMA 2017 is a big success and we hope that we all enjoy these days.

Best regards,

Isabel Barbancho
Lorenzo J. Tardón
FMA 2017 General Chairs

FMA 2017 Organizing Committee

General Chairs

Isabel Barbancho (Universidad de Málaga, Spain)
Lorenzo J. Tardón (Universidad de Málaga, Spain)

Local Committee

Ana M. Barbancho (Universidad de Málaga, Spain)
Alberto Peinado (Universidad de Málaga, Spain)
José L. Santacruz (Universidad de Málaga, Spain)
Marcelo Caetano (Universidad de Málaga, Spain)
Alejandro Villena (Universidad de Málaga, Spain)

Program Committee

Olivier Adam (UPMC, Paris, France)
Isabel Barbancho (Universidad de Málaga, Spain)
Filippo Bonini Baraldi (INET-md, FCSH, Universidade Nova of Lisbon)
Pierre Beauguitte (DIT, Dublin, Ireland)
Emmanouil Benetos (Queen Mary University, London, UK)
Emilios Cambouropoulos (Aristotle University of Thessaloniki, Greece)
Dorian Cazau (ENSTA Bretagne, Brest, France)
Darrell Conklin (University of the Basque Country UPV/EHU, Donostia-San Sebastián, Spain)
Carroll Dave (Dublin Institute of Technology, Ireland)
Thomas Fillon (Parisson, Paris, France)
Emilia Gomez (Universitat Pompeu Fabra, Barcelona, Spain)
Andre Holzapfel (Royal Institute of Technology, KTH, Sweden)
Matija Marolt (University of Ljubljana, Slovenia)
François Picard (Paris-Sorbonne University, France)
Aggelos Pikrakis (University of Piraeus, Greece)
Julien Pinquier (IRIT, Toulouse, France)
Lorenzo J. Tardon (Universidad de Málaga, Spain)
Peter Van Kranenburg (Meertens Institute)
Anja Volk (Utrecht University, The Netherlands)
Chris Walshaw (University of Greenwich)
Tillman Weyde (City University London, UK)

FMA 2017 Program at a glance

	Wednesday June 14	Thursday June 15	Friday June 16
9:30	Registration	Opening of the day	Opening of the day
10:00		Tutorial	Keynote Talk 2
10:30	Opening of FMA 2017		
11:00	Keynote Talk 1	Coffee Break	Presentation of Posters & Demos
11:30		Oral Session 3	Coffee Break
12:00	Coffee Break		Poster & Demo Session 2
12:30	Oral Session 1		
13:00		Presentation of Posters & Demos	
13:30		Lunch Break	Best Paper Award. FMA 2017 Closing
14:00	Lunch Break		
14:30			
15:00		Oral Session 4	
15:30	Oral Session 2		
16:00		Poster & Demo Session 1	
16:30	End of Scientific Program		
17:00		End of Scientific Program	
18:30	Walk in the city		
20:00	Malaga Folklore & Cocktail at Rectorado		
20:30		Gala dinner	
21:00			
22:00			

FMA 2017 Detailed Program

Wednesday, June 14	
09:30	Registration
10:30	Opening of FMA 2017
11:00	**Keynote Talk by Prof. Dr. Lars-Christian Koch** **Models of oral transmission of music - Permutation as a basic concept of Raga elaboration in North Indian Music**
12:00	Coffee Break
12:30	**Oral Session 1** **Analysis of Folk Music Styles** **Session Chair: Frank Scherbaum** • ANALYSIS OF THE TBILISI STATE CONSERVATORY RECORDINGS OF ARTEM ERKOMAISHVILI IN 1966, Frank Scherbaum, Meinard Müller and Sebastian Rosenzweig • INTERVAL STRUCTURES IN GNAWA MUSIC: DYNAMICS OF CHANGE AND IDENTITY, Lorenzo Vanelli • RHYTHMIC PATTERNS IN RAGTIME AND JAZZ, Daphne Odekerken, Anja Volk and Hendrik Vincent Koops • SCALES IN LITHUANIAN TRADITIONAL FIDDLING: AN ACOUSTICAL STUDY, Rytis Ambrazevicius • MICROINTERVAL MODALITY IN IRISH TRADITIONAL MUSIC – AN EMPIRICAL APPROACH, Ryan Molloy
14:00	Lunch Break
15:30	**Oral Session 2** **Music Composition and Synthesis** **Session Chair: Pierre Beauguitte** • AN IMMUNE-INSPIRED COMPOSITIONAL TOOL FOR COMPUTER-AIDED MUSICAL ORCHESTRATION, Marcelo Caetano, Isabel Barbancho and Lorenzo Tardón • A SEMI-AUTOMATIC METHOD TO PRODUCE SINGABLE MELODIES FOR THE LOST CHANT OF THE MOZARABIC RITE, Geert Maessen and Peter Van Kranenburg • SYNTHESIS OF TURKISH MAKAM MUSIC SCORES USING AN ADAPTIVE TUNING APPROACH, Hasan Sercan Atlı, Sertan Şentürk, Barış Bozkurt and Xavier Serra
16:30	End of June 14 Scientific Program
18:30	Walk in the city
20:00 22:00	Malaga Folklore & Cocktail at Rectorado

Thursday, June 15	
9:30	Opening of the day
10:00	**Tutorial. The Persian Musical System and the Dastgàh Recognition** **Peyman Heydarian**
11:00	Coffee Break
11:30	**Oral Session 3** **Session Chair: Chris Walshaw** **Automatic Extraction of Information from Audio and Video Music Files** • IMPROVED ONSET DETECTION FOR TRADITIONAL IRISH FLUTE RECORDINGS USING CONVOLUTIONAL NEURAL NETWORKS, Islah Ali-Maclachlan, Carl Southall, Maciej Tomczak and Jason Hockman • TUNE CLASSIFICATION USING MULTILEVEL RECURSIVE LOCAL ALIGNMENT ALGORITHMS, Chris Walshaw • IMAGE-BASED SINGER IDENTIFICATION IN FLAMENCO VIDEOS, Nadine Kroher, Aggelos Pikrakis and José-Miguel Díaz-Báñez • DEMYSTIFYING FLOWS BASED ON TRANSITIONAL PROPERTIES OF IMPROVISATION IN HINDUSTANI MUSIC, Achyuth Samudrala and Navjyoti Singh
13:00	Short presentation of the posters
13:30	Lunch Break
15:00	**Oral Session 4** **Score Based Methods and Educational Approaches** **Session Chair: Marcelo Caetano** • SOURCE SEPARATION BY SCORE SYNTHESIS IN SPANISH FOLK MUSIC, Sergio P. Paqué, Marcelo Caetano, José L. Santacruz, Lorenzo J. Tardón and Isabel Barbancho • SCORE-INFORMED SYLLABLE SEGMENTATION FOR JINGJU A CAPPELLA SINGING VOICE WITH MEL-FREQUENCY INTENSITY PROFILES, Rong Gong, Nicolas Obin, Georgi Dzhambazov and Xavier Serra • WHAT IF FIDDLING SOLVES YOUR PROBLEMS?, Enric Guaus, Oriol Saña and Laura Ramos
16:00	**Poster & Demo Session 1** **Session Chair: Lorenzo J. Tardón** • MELODIC PATTERN CROSS-OCCURRENCES BETWEEN GUITAR FALSETAS AND SINGING VOICE IN FLAMENCO MUSIC, Inmaculada Morales, Nadine Kroher, and José-Miguel Díaz-Báñez • EXTRACTION AND CLASSIFICATION OF ORNAMENTATION IN FLAMENCO SINGING: AN EVOLUTION-BASED APPROACH, Inmaculada Marqués, Nadine Kroher, Joaquin Mora and José-Miguel Díaz-Báñez • THE AEPEM COLLECTION: A SET OF ANNOTATED, Pierre Beauguitte • DEMONSTRATION OF SOFTWARE FOR CHORD DETECTION AND ANALYSIS, Eva Ferkova and Michal Sukola
17:00	End of June 15 Scientific Program
20:30	Gala dinner

Friday, June 16	
9:30	Opening of the day
10:00	**Keynote Talk by Dr. Dolores Vargas Jiménez** **Musical interpretation through the iconography of the dance in Pablo Ruiz Picasso**
11:00	**Short presentation of the posters and demos**
11:30	Coffee Break
12:00	**Poster & Demo Session 2** **Session Chair: Ana M. Barbancho** • ADAPTATIONS OF SLOVAK FOLK SONGS FOR PIANO IN THE CONTEXT OF (ETHNO)MUSICOLOGICAL ANALYSIS, Eva Ferkova and Hana Urbancova • CROSSED-EYED CHORO: FORMAL DEFORMATIONS IN BRAZILIAN CHORO, Cibele Palopoli • METER IN “ESPERANDO NA JANELA” (2000): A GLIMPSE INTO HYPERMETRIC SHIFTS AND PERCEPTION, Eduardo Sola Chagas Lima • RASPBERRY PI + LEGO = FLAMENCO DEMONSTRATOR, Alejandro Villena, Joaquín Cáceres, Marcelo Caetano, Isabel Barbancho and Lorenzo Tardón
14:00	Best Paper Award. FMA 2017 Closing

Contents

Keynotes

Tutorial

Oral Session 1

Oral Session 2

Oral Session 3

Oral Session 4

Poster Session 1

Poster Session 2

ISBN: 978-84-697-2303-6

Keynotes

Folk Music Analysis
MÁLAGA, Spain 14-16 June 2017

Keynote 1

Prof. Dr. Lars-Christian Koch is Head of Department of Ethnomusicology and Berlin Phonogram Archive at the Museum of Ethnology in Berlin (Germany) and Professor for Ethnomusicology at the University of Cologne and Honorary Professor for Ethnomusicology at the University of the Arts in Berlin. He was Guest Professor at the University of Vienna and at the University of Chicago. He has conducted field work in India, as well as in South Korea. His research focuses on the theory and practise of North-Indian Raga-Music, organology with special focus on instrument manufacturing, Buddhist music, popular music and urban culture, historical recordings, and music archaeology.

Models of oral transmission of music - Permutation as a basic concept of Raga elaboration in North Indian Music

Oral transmission of musical content are cultural strategies in diverse forms. Especially on the Indian Subcontinent, these strategies are obvious as during the elaboration of a Raga in North Indian Music the century old concept of permutation plays a central role it affects the melodic structure of a raga resulting in a different melodic concept compared to Western Music. Some analysis of selected melodic patterns from Indian raga music will illustrate these differences.

Furthermore, permutation is one of the main aspects of teaching North Indian Music, which generates a wide vocabulary of patterns as a repository for improvisation within a set melodic framework. By means of case studies of selected raga-s this lecture should illustrated how the ability to improvise is taught in a traditional way and how this art is applied in performance.

Keynote 2

Dr. Dolores Vargas Jiménez (Flamenco dancer) Dolores Vargas Jiménez was born in Málaga. She received the B.E. degree in Geography and History and her Ph.D. degree from the University of Málaga, Spain. Her Ph.D. Thesis "Picasso: Iconography of Dancing" was awarded the "Premio Málaga de Investigación" by the Academies "Bellas Artes de San Telmo" and "Malagueña de Ciencias" in 2013. Now, she teaches history at the "Escuela Superior de Turismo Costa del Sol de Málaga". She combines her academic facet with flamenco dancing. She has travelled around Spain, Canada, Sweden and France showing the flamenco dancing. In some of her performances, she is accompanied by her sister Mercedes.

Musical interpretation through the iconography of the dance in Pablo Ruiz Picasso

In the artistic production of Pablo Ruiz Picasso there are multiple examples of how popular music, translated in dance poses, features in the works of the Malaga artist. The images of dance in Picasso are the result of his emotional and sensorial experience while hearing the music, and his contemplation of the dance.

MUSICAL INTERPRETATION THROUGH THE ICONOGRAPHY OF THE DANCE IN PABLO RUIZ PICASSO

Dolores Vargas Jiménez, Ana María Barbancho Pérez
Universidad de Málaga, ATIC Research Group, ETSI Telecomunicación,
Dpt. Ingeniería de Comunicaciones, Campus Teatinos, 29071 Málaga, Spain
lolivargas@hotmail.com, abp@ic.uma.es

ABSTRACT

Popular, traditional or folk music has been present since its origins in the world of the History of Art. The different artistic representations, understood as the manifestations of the spirit, collect and capture sequences of the vibration and impact of music interpreted by dance, which was considered the first of the arts, just as the art of music was born with the human being .

In the artistic production of Pablo Ruiz Picasso there are multiple examples of how popular music, translated in dance poses, features in the works of the Malaga artist. The images of dance in Picasso are the result of his emotional and sensorial experience which he perceived listening to the popular music and contemplaty the dance.

1. INTRODUCTION

To speak of the artist, considered the genius of the Art of the XX century, and try not to reiterate is quite complicated, since we could practically say that his life and his work have been completely historyd. Many reflections have been made about the Picassian world, since almost everything has been scrupulously reviewed not only by art historians and plastic artists, but also by psychologists, writers and musicians.

Pablo Ruiz Picasso was a creator of great charisma and personality. After the study of his production, new approaches arise, a different reading of the artistic medium.

But what is really interesting is that before being in contact with the world of dance, it was a theme that appeared continuously in his work from his earliest childhood. Even after divorcy his wife, a Russian dancer and the way of life that followed in this time, he continued to resort to this inexhaustible source of sensations,. He continued to surround himself with musics and dancers. In addition their parties and meetings, in one way or another, were always brought to live with dancers.

The encounters with artists from other disciplines brought him inspiration, as if part of that internal force was transferred to him to be able to contemplate it and later, thanks to his retention capacity, he transferred this to his work.

He drew first everything that he was interested in, although later or not he may not have taken it to his canvas or other supports to shope. In all the artistic techniques that he carried out we found the influence of the dance movement and of course with his own music.

Although Picasso was also encouraged to actively participate in the various improvised dances around him, we thought that the greatest enjoyment was experienced through the visualization of this art. Experts in psychology point out the importance of the environment that surrounds a person the first few years of his life. The retentive capacity is in full development and the child absorbs everything he/she sees, touches, smells, stays in his/hers subconscious forever.

The dance has fascinated artists of all guilds, but in particular for painters it has been a way of identification, an instrument to be able to communicate the movement in their creations, to create a space full of rhythm and sensuality without limits. Kandinsky (1999) duelved into the importance and complementarity of the arts, the inspiration of one art in another only succeeds if the inspiration is not external but of principle. Each art possesses its forces, which can not be replaced by those of another art.

Throughout history, dance never ceased to be art but has not always been recognized and integrated within the classification of the Arts. From the XV century it is cataloged as an elegant and pleasant art, next to Painting, Sculpture, Architecture, Music, Poetry, Theater and Dance, all of them separated from Crafts and Sciences. Already in the sixteenth century, Francesco from Holland, coined the term of Fine Arts, referring to Visual Arts. In the seventeenth century in the treatise on Architecture of François Blondel will include Architecture, Poetry, Eloquence, Comedy, Painting and Sculpture, later adding to Music and Dance. Discovering that all the Arts have a factor in common: all are a source of pleasure

ISBN: 978-84-697-2303-6

for those who experience them, sharing the idea of beauty. But it will not be until the 18th century when Batteaux presents a list of Fine Arts: Painting, Sculpture, Music, Poetry and Dance, to be more exact, art of movement, L'art du geste, to which he will add Architecture and Eloquence. He will be the first to consider the Fine Arts as mimetic, an idea that his contemporaries did not likes much, it did not enjoy popularity but it was quickly accepted.

Already in the nineteenth century two of them, Music and Dance, will be separated, leaving three visual arts: Architecture, Sculpture and Painting. Mimesis will keep together Music, Dance and Mimicry, its function being to externalize the deepest feelings, those of the soul. With dance there appears a flowering of rhythmic impulses from which we can obtain great amount of sensations. Sometimes gestures can be more expressive than multiple words. Françoise Delsarte (1811-1871) found after his research that every emotion or brain image corresponds to a movement or at least an attempt to movement.

Perez Rojas (1994) pointed out how the musical feeling is also translated in the serpentine movements of the figures. A soft music produced by the whisper of nature or the chord of a hidden violin which invites the dance, this will be translated by the artists sensitive to it. There are many examples of Spanish artists that will vibrate with this interpretation.

For Picasso the whole interest of the Art is in the beginning, after the beginning comes the end. The celebrations carried out by the Málaga society of the time in the years of his childhood and the vacacionales stays of his youth, will be engraved in his eyes. What we perceive influences when expressing ourselves, it is the reaction to the perception translated in the pleasure of the contemplation. The estimation of artistic values is relative and depends on education and the socio-cultural environment that surrounds us to a great extent.

Below we show two figures where the musical instruments and some elements together with the different sections of Flamenco are present in the different artistic stages of Picasso.

Following the biography of the artist, in the following six sections we will see how popular music is present during his life, and thus is reflected in his work.

In the section *Everything begins in Malaga*, we will explore his beginnings. In *the Cafes Cantantes* we will analyze some works inspired by these very famous places in Spain. During *his first visit to Paris*, new music is discovered by Picasso. With *the origin of Modern Art*, the artist turns his eyes towards the exotic and the known, in the same way that is inspired by the traditional to *decorate the College of Architects of Cataluña*. The women who had a presence in his life, were also shaped in their compositions as well as the women of the entertainment world dedicated to music and dance.

2. EVERYTHING STARTS IN MÁLAGA

Pablo Ruiz Picasso was born in Malaga, on October 25, 1881 in the current Plaza de la Merced. The Malaga society was going through difficult times from the economic point of view. The pictorial art will accompany Pablo Ruiz Picasso from his birth. His father, as well as a painter, also worked as a teacher of Drawing at the School of Arts and Trades of San Telmo, being named in 1879 as a curator of the Municipal Museum, founded by the painter Muñoz Degrain. Its first ten years of life, fundamental in the formation of a person, will take place in Malaga.

Experts in psychology point out the importance of the environment that surrounds a person the first few years of his life. The retentive capacity is in full development and the child absorbs everything he sees, touches, smells, stays in his subconscious forever. The experiences in his hometown that he remembered most perfectly, were not images of childhood fleeting. Even if he stopped living in it at almost ten years, every summer or holiday period until he was nineteen he used to return.

From an early age he would have a direct experience with the festive events developed in these years. We thought that he would see dance in each of the celebrations celebrated in his native city of the time: in verbenas, fairs, May crosses, celebration of Corpus Christi, flamenco shows in the singing cafes (although the entrance to the children to these places Was allowed). We could say that in each of these celebrations music and dance were indispensable protagonists. From his first drawings, action and movement are already present in his works based on perception, agglutinating elements and using codes that awaken feelings.

Picasso prided himself on the fact that in his childhood he used to frequent and interact with the gypsies who lived in the area adjoining the Muslim fortress of the Alcazaba in Málaga, in the so-called Chupa and Tira or Mundo Nuevo neighborhoods. The artist told how they had taught him to love cante jondo and to dance flamenco in a rudimentary way. The gypsies taught me many tricks, he used to say mysteriously (Richardson 1995). The world of flamenco will fascinate you throughout your life, felt something special when listening to the sounds of the guitar or enjoy the contemplation of the majesty of the dance.

He was proud of that inheritance acquired during his childhood. He experienced an intense emotion for the world of bulls and a great sensitivity for "lo jondo". The deep, the inner. It is the peculiar adjective to describe the feeling in flamenco; To those letters that show life, to sing in grief and in joy: what is lost is sung, as Arrebola (1986) collects in the singular way of externalizing art in the world of flamenco.

As a child he would be a participant in the festivals and cultural events organized in the Lyceum of Malaga. Conferences, concerts, floral games, recitals and dances took place in the cultural life of Málaga at the time. From his hometown he would begin to absorb everything related to the artistic world.

The economic situation of the family Ruiz Picasso became more and more complicated. Jose Ruiz decides to change his place of professor in the School of Malaga by another like professor of drawing in the institute La Guarda of La Coruña. The change will be radical. Picasso's younger sister, Conchita, fell ill with diphtheria, dying on January 10, 1895 at the age of seven. This dramatic event will mark the whole family that will live in sadness. Two months later, he will accept the vacancy of professor of drawing in the Lonja of Barcelona. Picasso recreates a typical local folk scene in 1895 and Woman with tambourine, Picasso's parents showed great interest in the artistic and cultural world of his time, as well as being great fans of the show business.

The first major works that he will develop in Barcelona will be based on the religious theme, very tasteful of the time and very appropriate for a young man who begins with his pictorial work. In the first years of his stay in the city of Barcelona there are already a series of dancers combined with other scenes. In 1896 he began to draw interior scenes with dancers accompanied by a guitarist or by a pianist perfectly located at the foot of the stage, in a lower plane than the dancers. As an example, Café de Chinitas, a famous café in Malaga, which Picasso could possibly attend in his holidays in his hometown.

These types of scenes were responsible for making the customs and way of life known in a very peculiar way in the south of Spain, where the festive party of singing and dancing was renamed juerga flamenca. But not only by foreign artists but also by Spanish.

In the period of 1897-1898 he become student of the Royal Academy of Fine Arts of San Fernando of Madrid. There are not too many preserved works of this period. The atmosphere did not accompany him too much, a very hard winter, loneliness and contracting scarlet fever will diminish the restless spirit of the artist. But in the midst of all this, a scene starred by two Flemish bailaoras (picture 17) arranged in space in a singular way. In the lower left corner is occupied by the flamenco painting, that is to say by the accompaniment of the bailaoras. Three smiling female profiles seated very well defined direct their eyes towards the front where the bailaoras should be, that in this case occupy the superior zone of the paper. Usually in the cafes concerts the performances were represented on the tablao or stage, remarkably elevated of the public for its optimal visioning. Together with the women sitting in the same position the guitarist from which we can observe his saddened and tired countenance. Beside him, another female figure with his face partially hidden by the guitar's neck.

To recover from his poor health, Picasso decides to accompany his friend Pallares to his hometown, Horta de Ebro. There he will work in the field performing the different agricultural tasks that would shape his notes, drawings and oils. Sample of it is the composition Customs Aragonesas that would be presented in the National Exhibition of Bellas Artes of Madrid of that same year, obtaining an honorific mention. Once again pick up the essence of the place through the popular music of the locality.

In the drawing scene of tablao flamenco, a woman stands gaudily seated (the body facing the spectator and the head in profile), left hand in the waist and right on the leg, rising on outlined lines as a tablao. Along with her Picasso also draws a concentrated guitarist. On this same sheet and on the left side, another woman standing wrapped in her shawl repeats the position of her face shown equally in profile.

3. THE “CAFÉS CANTANTES”

From 1896 onwards, there were successive drawings inspired by the interiors of the cafes cantantes where you could enjoy flamenco together with a variety show. In Barcelona the tradition of this type of establishments begins to develop in the middle of century XIX. These premises were responsible for making flamenco known to the public, not only in Andalusia, where the most famous were found, they also proliferated in other places such as Madrid and Barcelona. They acquired great fame from the middle of the 19th century until the beginning of the 20th century. This type of premises had a series of characteristics that were very attractive to the bohemian artists of the time. There they could enjoy flamenco art along with other demonstrations, while the attendants could drink and talk animatedly.

We find an interesting drawing by Picasso titled Café Concierto created in Barcelona in 1900, where the Malaga artist collects a scene typical of the singing coffees of the time. Once again the main character of the composition is a dancer in full performance. The vigorosidad of the moment is reinforced by the figure of the expressive cantaor that accompanies the sequence with the sounds of his voice. The festive atmosphere and fun is perfectly captured by the other characters that make up the scene. These are distributed between the side boxes, divided into two floors and those that occupy the tables distributed at the foot of the tablao, focusing their attention on the stage.

We have found it attractive to compare the composition of Picasso with a work by the Catalan painter José Llovera and Bofill entitled Baile Flamenco made in 1890. If we look at the environment, we could say that both images collect the same place. There are some similarities that we can appreciate, such as the distribution of space, the side boxes of two floors with similar decoration and even the three points of light on both sides of the scene. Picasso has neglected some details of the setting of the stage, although his work was done ten years after the one made by Llovera and Bofill and possibly, if it is the same place, it could have undergone some transformation. But in both illustrations the essence of the place has been collected. In the singing cafés flamenco coexisted with bowling school dances (totally academic), with numbers of magic and singers of different styles. There is no doubt that the aesthetic and plastic strength of a flamenco painting on stage has attracted the attention of numerous artists, not only Spaniards but foreigners seduced by the exotic and picturesque sequence

The bailaora, both in the composition of Bofill and Picasso, is touched in a wide-brimmed hat, wrapped in her stylized figure in a Manila shawl and possibly in a tail coat, although in the Picasso drawing we do not appreciate to distinguish it with accuracy. We could suggest that both were playing garrotín, a very popular dance at the end of the 19th century. It is a flamenco club of uncertain origin, considered as a genre of import, which takes most of its musical elements from flamenco tangos of festive air, such as Tangos del Camino de Granada. This dance became very popular at the end of the 19th century, popularizing itself in Catalonia. Its interpretation will be lavished by the leading tablaos of the country and by the highest representatives of flamenco singing and dancing. It used to be customary for the bailaora to be accompanied by a hat, (currently the use of this complement is maintained) to interpret the different steps of the choreography, as an addition to the verses of the same that say:

Ask my hat,
My hat will tell you,
The little nights I spent
And the light that gives me,
The garrotín, the garrotán
Of the vera, vera, vera of San Juan.

4. FIRST VISIT TO PARIS

The French capital became the European cultural center at the end of the nineteenth century, increasing its popularity by the different Universal Exhibitions that took place there. The Universal Exhibition held in 1900 had great repercussions in the intellectual circles of Catalonia. Picasso sent a work called Last moments, being selected and exposed in the Spanish section of the decennial of the Grand Palais, dedicated to the Spanish pavilion.

The taste for the shows would be filled in Paris. He made a drawing where the development of a belly dance show can be appreciated, as Picasso notes along with his signature. It must have been very new to be able to visualize a show of these characteristics, possibly unknown to the artist up to those moments, hence the choice of the iconographic motif for the letter. Oriental dances, French Cancán, Flamenco and festive dances are shown before the young Spanish artist.

5. THE ORIGIN OF MODERN ART

When we did Cubism, we had no intention of making Cubism, but of expressing what we felt (Elgar 1958). With this statement Picasso, not very fond of having to give explanations, showed the procedure followed to lead to this new artistic style. Under this intimate and personal criterion an important number of compositions were born.

In them we can find and discover a special movement. The figures float in space just as dancers do on stage. The rhythm is essential in all Picassian artistic creation, both pictorial and literary in which he used words, figures, musical notes, poems in a loop, the beginning is the end, as if it were a letter dance. Of difficult reading, the rhythm reappears here as well and of course is strongly present in cubism.

At first sight it seems to us that the works of this period are totally abstract, but this is not the case, since Picasso fled from abstraction. Brassaï (2002) explains how Picasso's painting is made of negotiations and eliminations, of ellipsis, of ruptures of forms, seems to be born frequently of free invention. But even when it seems to be far from reality, and even when his work covers every aspect of the fantastic or the surreal, its basis is a solid realism. The artist always gives us clues so we can recognize what is represented. The sensations that can suggest us the observation of a work of art lie in knowing what we are seeing. Gombrich (1987) mentions how pleasure derives from recognition. But not only in this, but also in the affinity that we can maintain with the represented, creating a mental union, an identification with the shaped.

The Demoiselles of Avignon is about the composition that more previous drawings realized. Multiple influences we find where the oriental dance and the dancer Mata Hari, with their exciting poses, acquire all the protagonism. Flamenco and more specifically, the Spanish guitar is very present during this stage of the artist.

6. THE COLLEGUE OF ARCHITECTS OF CATALUÑA

In one of the conversations that Brassaï maintains with Picasso he explains how in Barcelona he had a great impression. That sour, bittersweet song. That square full of girls, young people. Bags and jackets piled on the floor, and around each heap a circle of dancers and ballerinas rippling. It was so unforeseen. And the seriousness of the faces, tense, almost pathetic. Not a burst of laughter, not a smile. All solemn. I thought I was witnessing a religious ceremony.

For Picasso this folkloric dance was more than a mere amusement, all people are on the same level, regardless of social scale. Catalan sardana is nothing more than a collective dance whose meaning of ritual dance of the sun was forgotten over the centuries in its long evolution until reaching the Middle Ages, which is as far as our earliest news about it. That is why it has been tried to find in this dance a symbolism in its eight bars called "short" and sixteen "long", as a representation of the twenty-four hours of the day.

The first symbolize, with their melancholy intonation, the dark hours of the night, which the dancers execute barely moving; But afterwards the sixteen long ones, with their intonation of happy and luminous hours of the day, that the dancers symbolize dancing fiercely until the dance ends, which began with the notes of the call of the caramillo, like the song of the cock at midnight. Although the dances of corro were very habitual, of which there are more than one hundred and thirty only in the province of Malaga, Picasso related this form of dancing with the sardana, typical dance of the Catalan folklore.

In October of 1960 Picasso realized the drawings that decorated the facade of the College of Architects of Cataluña and Baleares. The technique used to pass the drawings to the concrete consisted in drawing these with sand blasting under pressure. Inside the building we can find two murals designed by Picasso, one of them collects a personal interpretation of a panoramic view of the city of Barcelona, called the mural of the arches and the other inspired by the sardana dance.

In the mural dedicated to the sardana can be identified, through some arches, the silhouette of the castle of Monjüit. At the bottom and next appears a kind of aqueduct, more static than the previous ones. In front the face presents a unique composition: two concentric sardanas, with the one of the children in the interior. The architect Javier Busquets was in charge of contacting Picasso and proposing the idea of decoration of the building. After several conversations he took to the artist graphic documents where he could refresh his childhood memories in Barcelona, through his parties like the fair of Santa Lucia and the procession of the Corpus with its giants and cabezudos. Picasso saw and remained silent. In the final work the prodigious visual memory of the artist was captured. Finally on October 18, 1960, Picasso called Busquets to inform him that the drawings were ready.

Years later Busquets commented that the simple lines in the friezes of the College of Architects of Catalonia and Baleares, located in the heart of Barcelona in front of the Cathedral, are a real effort of synthesis made by an artist of eighty years after a long search life Of the intimate forms of art.

7. PICASSO AND WOMEN-DANCERS

Picasso's fascination with the figure of the woman is evident throughout his extensive artistic production. There are several models of woman that he reflects in his work. But there is one that appears continuously, which will repeated constantly throughout his life: the image of the dancer. These representations highlight the passion of the performers when they dance. For Picasso contemplation of the dance was a source of inspiration, which nurtured his creativity. For him these women possessed on elf, a special force that he wanted to translate into his works thanks to the movements of his pencil, pen or brush. Through his drawings we perceive sensorial, aesthetic, iconographic and technical aspects, by which we enter into the artist's intimacy.

He was able to choose this feminine type for the sensual load that he gave off, but also for the bohemian life in which they could be found and at the same time, for the image of an easy woman with whom to establish intimate relationships, although in fact it was not so. Not only Picasso will look for certain representations. Matisse resorts to representations of odalisques.

The myth personified in Olga collapses, but the woman who dances will continue to star in his works until the end of his life. And although his next partners will no longer be entertainers, the master will like to portray them, giving them in their paintings an artistic halo. For Picasso dance and dance are the movements of Flamenco Art, so rooted in his character, producing feelings that no other type of rhythm will bring.

Statements of the artist himself revealed that for him, music was the pasodoble and strumming of a flamenco guitar. And dancing is those senses swaggering with hips, arms to the sky and the rhythmic movement of the feet. There are many dancers and dancers who draw, paint and record in all their trajectory. Some of them we can identify them with famous artists of the time, since in Paris pass the best stars of the dance of the moment. Even during his first exhibition at the Vollard Gallery in Paris, he was known as the painter of dancers, in the words of critic Gustave Coquiot (Daix 1989).

Several were the women who passed through the life of the artist, those who were important appear reflected in his work, with whom he shared his life draws or paints them in a dance attitude. All his women in one way or another immortalize them by dancing. Picasso liked to represent them wrapped in rhythm and movement, loaded with sensuality that we can interpret as a dance.

7.1. Sentimental partners

A new woman appears and is born an unpublished form of her art, an unforeseen mode of expression (Cabanne 1982).

Rosita del Oro, an acrobat of the circus of Tivoli from 1897 to 1900, is the first relationship to exist. It will be thanks to Rosita that Picasso knows and is fascinated by the circus world (Matabosch 2006).

Fernande Olivier, 1904-1911, the drawing Fernande dansant appears next to the artist Paco Durrio who used to play the flamenco guitar, Picasso said that he played for malagueñas (Richardson 1995).

Eva Gouel, begins the relationship with Picasso in 1911 until 1915 that she passes away.

Gaby Depeyre comes to the artist's life in 1915, she sang and danced in a cabaret in Paris

Irene Lagut maintained a relationship with Picasso during the years 1916 and 1917. She worked at a Music Hall in Paris.

With the arrival in 1917 of **Olga KoKkova** and after several failed relationships, the artist marries this Russian dancer. He never painted her dancing. They had a son, Paul.

In 1927 he met young **Marie Therese Walter**, she was 17 years old, Picasso 50. Born of this relationship was Maya in 1935. Both portrays are of her dancing.

Dora Maar met Picasso in 1936, she was a woman ahead of her time: photographer, painter, sculptor and poet. She photographed the process of Guernica.

In 1943, Picasso meets the young painter **Françoise Gilot**. From this relationship were born two children: Claude and Paloma. Françoise is the protagonist of his production, example of, it is the composition the joy of living of 1946. She leaves him in 1953.

Picasso met **Jacqueline Roche** in 1952, married her in 1961, after being widowed by Olga. He shared with him the last twenty years, she was flooding his work. The Spaniard portrayed her like Lola de Valencia, famous Spanish dancer portrayed by Manet.

7.2. Dancers in their strokes.

The dancer, (Starobinski 2007) assumes an illusory role, represents a flower, a bird, a divinity or a personage of the cultural tradition. For the artists, the dancers have been a source of expiration. Before the retinas of Picasso the muses of the dance wandered, immortalizing them with their strokes.

Here are some of them:

The Bella Chelito, stripper dancer of the time dazzled the artist when he played the flea in 1902. Picasso remembered his whole life this song.

Sada Yakko, Japanese dancer to whom Picasso made an advertising poster and several drawings in 1900.

La Nana or Enana, made a portrait in 1901, just as it had been preceded by painters painting the famous Spanish dancers of European and American fame.

Jeanne Bloch, famous dancer of cancan, 1900. Picasso was very attracted by this type of dance.

La Bella Otero, in 1901 the woman with jewels, inspired by it, performed the work. Famous Spanish dancer known worldwide.

Jane Avril, muse of the painter Toulouse Lautrec, was one of the most famous cancán dancers. Picasso portrayed it twice, in 1901 and 1902.

Mademoiselle Bresina, Spanish dancer, portrait of 1903.

Mademoiselle Leonide, 1910, gives us the feeling of floating. This drawing, which appears cubist and realistic elements appears in a book by Max Jacob.

Olga with a mantilla, although she never painted it dancing if she wanted to paint it like a typical Spanish woman in a mantilla in 1917, yes, an improvised mantilla, since it was the crochet rug that had the table of the hotel where they were staying. Picasso will again innovate with its blankets of the costumes of the Russian Ballets in 1919.

Blanquita Suárez Picasso painted it in 1917, after seeing her performing at the Tivoli Theater in Barcelona.

For Picasso, the figure of the dancer or bailaora will appear accompanied in his artistic production by the picador, 1960 series and by the harlequin, as the numerous examples show. In one of his last works, Picasso goes back to its origins, refers to the traditional folklore of Malaga, the Verdiales.

8. CONCLUSIONS

Artists receive influences from all the arts, being essential the integration of each other, to achieve an enriching synergy that results in an inexhaustible source of inspiration. In this Malaga artist the most diverse manifestations of art meet and converge, being an example impossible copy. For the artist, music and dancing meant fun, sexual union and passion, but above all, feeling.

We find this passion for dance in the interpretation he makes of the women he portrayed, although he was not inspired so much by the classical dancers as by those who interpreted other more vibrant styles for the artist: flamenco dancers, cancán dancers or poses of oriental dances. In this way the artist fled from the embedding of the academic gestures of classical dance. However she also collected snapshots of these classical ballet dancers during their moments of rest, where they really showed themselves as they were.

We are faced with a new vision of the work of Picasso. An interesting quote by the German philosopher Nietzsche (1998) explicitly illustrates the definition of the untiring quality of Picasso's creation: what distinguishes a genuine original head is not to be the first to see something new, but to see old things as new, seen all over the world and not taken into account by anyone.

A historian can approach the patterns and influences that an artist can receive and channel in his works, but he can not delve into the intimate experiences and the final decisions inserted in them.

The creation never had limits for the Málaga artist, wrapped in an eternal rhythm and movement, without ties or conventions, in a free, active and alive way. Pablo Ruiz Picasso always found inspiration loving, hating, painting and of course, dancing.

9. REFERENCES

ARREBOLA A.(1986). *El sentir flamenco en Falla y Picasso.* Universidad de Málaga.

BRASAÏ (2002). *Conversaciones con Picasso.* Madrid. Turner Ediciones.

BOURCIER P.(1981). *La danza en Occidente.* Barcelona :editorial Blume,

CABANNE P (1982) *El siglo de Picasso. El nacimiento del cubismo. Las metamorfosis, 1881-1937.* Madrid: Ministerio de Cultura.

DAIX P (1989). *Picasso creador. La vida íntima y la obra.* Buenos Aires: Editorial Atlántida.

ELGAR F.(1958). *Picasso, época cubista.* Barcelona: editorial Gustavo Gili.

GOMBRICH E.H.(1987) *La imagen y el ojo*. Madrid: Alianza Editorial.

GYENNES REMENYI, J.(1990). *Dalí, Miró, Picasso.* Madrid:editorial Eudema,

KANDISKY, V.(1999). *De lo espiritual en el arte.* Barcelona: Paidós Estética.

MATABOSCH G (2006) . "Rosita del Oro y Pablo Picasso, notas de un idilio barcelonés" en *Picasso y el Circo.* Barcelona: Museo Picasso.

NIETZSCHE, F. W.(1998). *Aforismos.* Madrid: editorial Edhasa,

PÉREZ ROJAS, J. y GARCÍA CASTELLÓN, M.(1994) *Introducción al Arte Español. El siglo XX, persistencias y rupturas.* Madrid: editorial Sílex.

RICHARDSON, J.(1995). *Picasso una biografía, vol. I, 1881-1906.* Madrid: Alianza Editorial.

RICHARDSON, J.(1997). *Picasso una biografía, vol. II, 1907-1917.* Madrid: Alianza Editorial.

RICHARDSON, J.(2007). *A life of Picasso. The triumphant years 1917-1932.* Nueva York: Alfred A. Knof.

SELDEN S. (1972). *La escena en acción.* Buenos Aires: Eudeba, editorial universitaria

STAROBINSKI J (2007). *Retrato de un artista como saltimbanqui.* Madrid. Adaba editores.

TATARKIEWICZ, W.(1987). *Historia de seis ideas, Arte, belleza, forma, creatividad, mimesis, experiencia estética.* Madrid: Editorial Tecnos.

VARGAS JIMÉNEZ, D. (2015) *Picasso: iconografías del baile.* Málaga: Centro de Ediciones de la Diputación de Málaga, CEDMA.

Tutorial

Folk Music Analysis
MÁLAGA, Spain 14-16 June 2017

Title: The Persian musical system and the dastgàh recognition

Peyman Heydarian, London Metropolitan University

Outline of the tutorial

A tutorial on Persian music analysis, covering the intervals; the *dastgàh* (the underlying system of Persian music); forms and composition; and the MIR methods for Persian dastgàh recognition.

The dastgàh, the underlying modal system of Iranian classical music, is a phenomenon similar to maqàm in Turkish and Arabic music. It usually represents the scale and tonic, and is to some extent an indication of the mood of a piece. Methods for computational identification of the tonic and scale in Persian audio musical signals will be presented. The feature sets, chroma (a simplified spectrum) and pitch histograms; the classifiers, Manhattan distance, dot-product, and bit-mask; and theoretical and data-driven templates will be presented and compared. Theoretical templates are constructed, either using the scale intervals or by making a note histogram of existing pieces. Data-driven templates are made by calculation of the chroma of available audio samples.

1.1 Persian Intervals

There are different views on Persian intervals [1, 2, 3]. Vaziri suggested a 24-tone equal temperament (24-TET), by analogy with the Western 12-TET scale. He defined *sori* (𝄲) and *koron* (𝄳) symbols to show half-sharp and half-flat quartertones, which are widely used in Iranian music [1]. However, in musical practice the quartertones are not fixed and, depending on the scale, the piece, or the performer's mood, they can be less or more than an equal quartertone. Farhat [2] suggests that in addition to the Western semitone scale, two intervals between a semitone and a whole (small and large variants of the three-quartertone), and an interval between a whole tone and a minor third (approximating to one and a quarter tones) should be recognised.

From a signal processing point of view, all we need to know that is that in addition to Western intervals, there are flexible quartertones in Persian music, which lay between two neighbouring notes a semitone apart, and that only a few of them are used in practice. The Persian repertoire can be played with 13 different notes: 7 diatonic notes, 3 semitones and 3 quartertones [2, 3]:

E F 𝄲F #F G #G A 𝄳B B C 𝄲C #C D

1.2 *Dastgàh*

Persian music is based on a modal system consisting of seven main modes and their five derivatives: *shur*, *abu'atà*, *bayàt-é tork*, *afshàri*, *dashti*; *homàyun*, *bayàt-é esfehàn*; *segàh*; *chàhàrgàh*; *màhur*; *ràst-panjgàh*; and *navà*. They fall into five different scale categories: *homàyun* and *bayàt-é esfehàn*, *chàhàrgàh*, *shur*, *màhur* and *segàh*. The scales are provided in [3]. Figure 1 shows the five principal scales, where 24-TET is assumed. Both fixed and moving accidentals are shown.

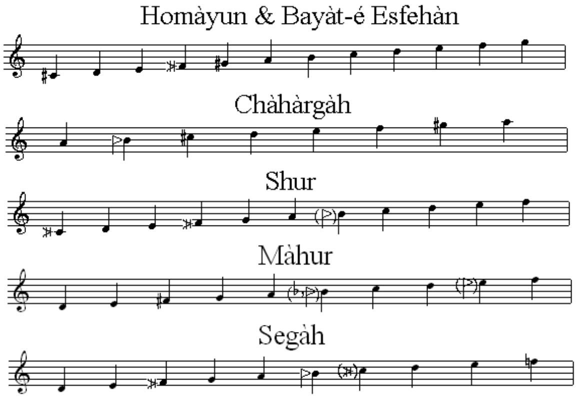


Figure 1: Scale intervals, based on 24-TET

2. DASTGAH ANALYSIS FLOWCHART

A *dastgàh* implies particular scalar intervals, a tonic, and modulations, and is to some extent an indication of the mood (emotional character) of a piece. The attributed emotions are usually culture-specific and depend on lyrics. A human listener recognises a *dastgàh* by one or more of these ways:

- Perceptually: based on the culture-specific mood of a piece
- Through melody/theme recognition: by matching the melody with known patterns
- Based on the intervals, the frequency of their occurrence, and order of the notes

The last two are clearer computationally. The bidirectional arrows between mode and melody (Figure 2) show that melody recognition reveals the mode and the mode can be used to improve melody recognition systems. A full *dastgàh* performance is recognised by tracking the modulations and the respective changing tonics.

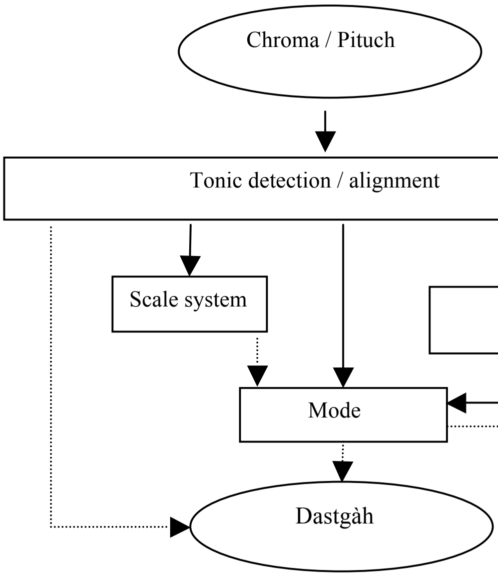


Figure 2: *Dastgàh* identification flowchart

Intended audience

Non-Western MIR researchers, ethnomusicologists, computational musicologists, Iranian and central Asian researchers

Peyman Heydarian's biography

Peyman Heydarian, born in Shiraz, Iran is an award-winning music scientist and santur virtuoso. Peyman started learning Persian music under the supervision of music masters, including Mojtaba Mirzadeh and Pashang Kamkar. His main instrument is the santur. He also plays the daf, piano, tar, violin, bouzouki, baqlama and harmonica. He has developed his own performance style on the santur and has adopted innovative tuning systems and techniques to play a multi-ethnic repertoire on the instrument.

Peyman has taught music and signal processing courses at different universities and has established and presided over a number of musical societies and bands, including the Music Association of Iranian Students (1998) and the National Iranian Students Orchestra (1999-2004). Since 1982, he has performed in Iran, USA, Canada, Syria, Jordan, Turkey, Greece, Italy, Hong Kong and the UK.

Peyman has been composing and recording music for films, including a BBC TV4 project "Axis of Light" and an Aljazeera TV film "Lover Boys". Peyman has been working in the field of music DSP since 1998. He holds BSc and MSc degrees in Electronic Engineering, from Shiraz University (1997) and Tarbiat Modarres University (2000) in Iran. And completed his MPhil on Signal processing at the Centre for Digital Music at Queen Mary, University of London (2008). Subsequently, he studied ethnomusicology at SOAS, University of London for a year (2010) and studied his PhD at London Metropolitan University (2016).

He is currently researching the possibilities of pushing the boundaries of the Persian music and santur performance; also developing algorithms for automatic recognition of the dastgàh / maqàm in audio musical signals.

REFERENCES

[1] Vaziri, A. N., *Dastur-e Tàr*, Tehran, 1913.

[2] Farhat, H., *The Dastgàh Concept in Persian Music*, Cambridge: Cambridge University Press, 1990.

[3] Heydarian, P. Reiss, J. D., "The Persian Music and the Santur Instrument", *Proceedings of the International Symposium on Music Information Retrieval*, London, 2005.

Oral session 1

Folk Music Analysis
MÁLAGA, Spain
14-16 June 2017

Analysis of the Tbilisi State Conservatory Recordings of Artem Erkomaishvili in 1966

Frank Scherbaum
University of Potsdam
fs@geo.uni-potsdam.de

Meinard Müller
International Audio Laboratories Erlangen
meinard.mueller@audiolabs-erlangen.de

Sebastian Rosenzweig
International Audio Laboratories Erlangen
sebastian.rosenzweig@audiolabs-erlangen.de

ABSTRACT

In this paper we try to obtain information regarding the musical thinking of Artem Erkomaishvili, one of the last master chanters of traditional Georgian chant. For this purpose, we analyse the recently determined F0-trajectories (Müller et al., 2017) for a set of chant recordings from 1966 in which Artem Erkomaishvili sang all three voices sequentially using two tape recorders in overdubbing mode. The purpose of our study is to determine the tuning of Artem Erkomaishvili's voice and how it compares to the models proposed by various researchers to reflect (in their opinion) the historical Georgian tuning. The analysis of the melodic pitch inventory shows that the sizes of melodic seconds sung by Artem Erkomaishvili vary over a range from approximately 140 to 240 cents, with a peak of the distribution at approximately 180 cents. We do not see evidence for any attempt to precisely use any particular or a small set of melodic interval sizes, as is suggested by some of the proposed tuning models. The harmonic analysis yields an interval distribution which is peaking at justly tuned fifths and octaves at 698 and 1203 cents, respectively. No observational evidence for stretched octaves, as suggested by some models, is seen. Analysing the joint pitch distribution, we find evidence for considerable voice interaction in which Artem Erkomaishvili maintained harmonic intervals despite considerable pitch fluctuations of the individual voices. In short, Artem Erkomaishivli's performance in 1966 seems to reflect a combination of strong harmonic and relaxed melodic thinking.

1. INTRODUCTION

Artem Erkomaishvili (1887-1967) is known today as a key representative of traditional Georgian singing of the 20th century and one of the last grand masters of Georgian chanting (*sruligalobelni*) (cf. Graham, 2015). In 1966, one year before his death, he was asked to perform all voices of a series of chants to save them for posteriority. His performance, part of which was recently remastered (Jgharkava, 2016), was recorded at the Tbilisi State Conservatory using two tape recorders, which were subsequently operated in what is now called overdubbing. The recordings were transcribed by Shugliashvili (2014). Although the use of the overdubbing setup originated from the lack of fellow chanters who could perform the repertoire, it turned into an advantage in view of an analysis of this data. Despite the fact hat polyphonic pitch analysis is still considered an enormous challenge in general situations, the sequential overdubbing considerably simplifies the task of determining the fundamental frequencies F0 (which for simplicity we will also refer to as pitches) for all voice segments. Details of the processing techniques can be found in Müller (2015). The corresponding time-stamped F0-trajectories have been made publicly available[1].

In the present paper, which is a direct follow-up study to Müller et al. (2017), we want to find out what we can learn from this unique set of recordings (respectively analysis results) regarding the characteristics of the tuning system(s) used by Artem Erkomaishvili. The topic of the authentic, historical Georgian tuning system has been a matter of intense and controversial discussion for a number of years, resulting in the proposition of several scale and/or tuning models which have little in common other than the untempered nature of the music (Erkvanidze, 2002; Gelzer, 2002; Westman, 2002; Gogotishvili, 2004 ; Kawai et al, 2010; Tsereteli and Veshapidze, 2014; Erkvanidze, 2016). Based on the analysis of recent field recordings in Svaneti/Georgia, Scherbaum (2016) took a conceptually different perspective on the issue of Georgian tuning systems. Since he found considerable differences in the sequential (melodic) and the concomitant (harmonic) intervals used by traditional singers, he concluded that a single scale/tuning model might not capture the complete tuning characteristics of Georgian vocal music. Instead, in line with Nikolsky (2015), he separately analysed the melodic and the harmonic pitch/interval inventory of the music. In the present study we go one step further and separately analyse the pitch organization in the recordings of Artem Erkomaishvili from a melodic, a harmonic and a voice interaction perspective.

The main part of our study is devoted to the attempt to use the time-stamped F0-trajectories of the individual voices in Artem Erkomaishvili's recordings to determine the associated melodic and harmonic pitch and interval inventories and to investigate how listening to prerecorded voices affected the tuning of Artem Erkomaishvili's singing. The results are discussed in the context of the predictions of the tuning models suggested by various researchers to reflect (in their opinion) the authentic, historical Georgian tuning practice(s). Our results suggest that voice interaction effects, evidence for which can clearly be seen in the recordings of Artem Erkomaishvili, should be included in the discussions of the tuning systems of traditional Georgian (and possibly other) vocal music. This might require a shift of attention from the purely melodic to the combined melodic-harmonic aspects of the music.

The paper is organized as follows. Following a brief recapitulation of the recording setup and the extraction of the F0-trajectories by Müller et al. (2017), we discuss the determination of the melodic aspects of the performance of Artem Erkomaishvili (Section 2.1). For the top voice segments we show that the individual F0-values, which make up the pitch tracks, exhibit a strong pitch clustering. We interpret the pitch values of the cluster centers (which we determine by k-means cluster analysis) to indicate the pitches of the notes of the mental melodic template Artem Erkomaishvili might have been using during his performance. From the pitch values of the cluster centers for the complete dataset, we determine the set of possible single-step melodic intervals for the complete performance. We compare (as a spot check) the properties of the resulting

[1] https://www.audiolabs-erlangen.de/resources/MIR/2017-GeorgianMusic-Erkomaishvili

ISBN: 978-84-697-2303-6

distribution with the results of a note analysis for a single chant using the Tony software (Mauch et al., 2014, 2015), and with the values of the single-step interval sizes from the predictions of some of the published tuning models for Georgian vocal music. Subsequently (Section 2.3), we discuss the harmonic aspects of the tuning used by Artem Erkomaishvili. In this context, we make use of the time-stamps for the individual voice segments determined by Müller et al. (2017) to estimate the F0-values for the concomitantly sung (harmonic) intervals. These also show a strong clustering, the properties of which we interpret to reflect the mental harmonic template Artem Erkomaishvili might have been using during his performance. As final aspect of our analysis, which according to our knowledge has previously been ignored in quantitative investigations of tuning in Georgian vocal music, we investigate (in Section 2.4) the joint pitch distribution of voice combinations for signatures of voice interactions. Considering a single chant, we find several instances in which Artem Erkomaishvili evidently maintained harmonic intervals despite considerable pitch fluctuations of the individual voices. Finally, in section 3, we conclude with a discussion of the main results of our study and their consequences for future work.

2. PITCH AND INTERVAL ANALYSIS

Figure 1 sketches the three-stage concept used during the recordings of Artem Erkomaishvili in 1966. In the first stage, only the lead (top) voice of a chant was recorded. In the second stage, Artem Erkomaishvili was singing the middle voice while listening to the recording of the lead voice. During the recording of the bass voice, he listened to the overdubbed recordings of the middle and top voice. The extraction of the F0-trajectories was also performed in a segmented way in that the extracted F0-trajectory for the first segment was used as constraint for the extraction of the F0-trajectory of the second segment, and so forth. For details of the analysis see Müller et al. (2017).

To make this audio collection better accessible for musicological research, one important task is to estimate the fundamental frequency (F0) trajectories of the sung pitches from the recordings using automated methods. While this is feasible with standard procedures in the case of monophonic music, the problem becomes much harder in the case of polyphonic music. In Müller et al. (2017), a graphical user interface (GUI) for semi-automatic estimation of F0 trajectories was introduced. The GUI allows a user to specify temporal-spectral constraint regions that guide the estimation process. Furthermore, the GUI provides visual and acoustic feedback mechanisms that can be used to control and refine the estimated results in an interactive fashion. In Müller et al. (2017), we applied this GUI for extracting the F0 trajectories of the sung pitches from the three-voice chant recordings performed by Artem Erkomaishvili. To this end, we first determined the recordings' structures based on the three-stage recording setup (see Figure 1). Subsequently, we determined the F0-trajectories for the lead, middle, and bass voices from the first, second, and third section, respectively. To this end, suitable visualization and sonification functionalities helped us in determining suitable constraint regions to guide the estimation process. All results, including the original recordings, figures of the visual representations, the estimated F0-trajectories, and the sonifications of these trajectories, have been made publicly available.[1] These results serve as an important basis for our subsequent analysis.

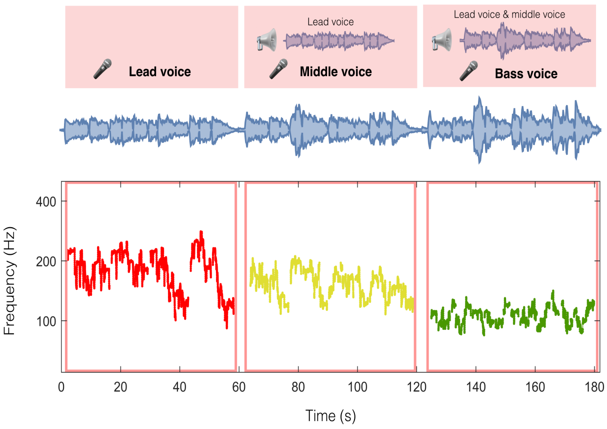


Figure 1. Sketch of the three-stage recording setup (top panel), the recorded waveforms (middle panel), and the F0-trajectories derived for the individual voices of chant no. 2 (Shugliashvili, 2014). The pink rectangles indicate the structure of the three-stage recording process.

2.1 Melodic Analysis

In the first step of our analysis, we determined the melodic pitch inventories of the lead (top) voice segments. These were always sung first and individually. We assume that, in case Artem Erkomaishvili believed that a particular chant should be performed in a specific scale, this will show up as a clustering of pitches around the intended "scale pitches" for this voice segment. This can be seen in Figure 2 for the chant *Aghdgomasa Shensa* (referred to by its chant ID no. 2 in Shugliashvili, 2014).

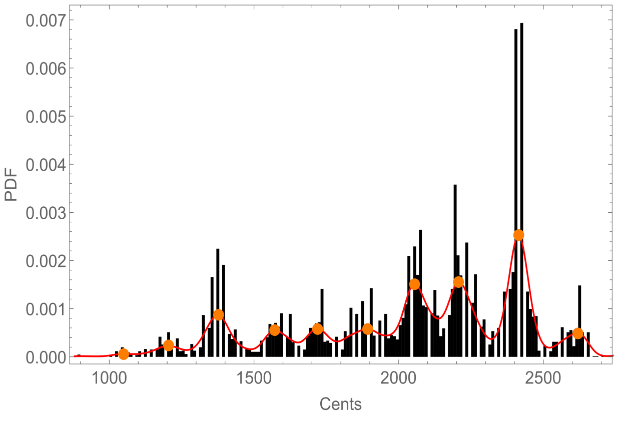


Figure 2. Pitch histogram (vertical axis scaled to match the sample PDF) and smooth kernel distribution (red solid line) of the F0-values in the top voice of chant no. 2. Note the clustering of the pitch samples. The reference note for all absolute cent calculations is A1 (55 Hz).

In order to determine what we assume to be the intended scale pitches quantitatively, we performed a formal cluster analysis (using the k-means algorithm) to determine the locations of the centers of the F0-clusters and the corresponding spreads. Figure 3 shows the resulting separa-

[1] https://www.audiolabs-erlangen.de/resources/MIR/2017-GeorgianMusic-Erkomaishvili

tion of the pitch set of the top voice of chant no. 2 into 11 pitch clusters.

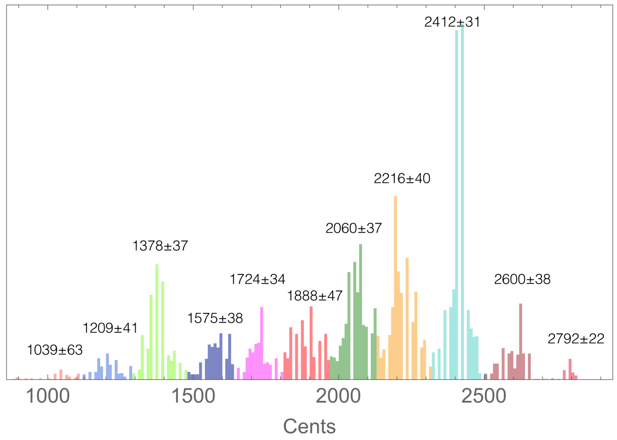


Figure 3. Pitch cluster histograms of the F0-values in the top voice of chant no. 2. The number on top of each cluster shows the cluster mean and the corresponding cluster standard deviation (in cents). The vertical axis is proportional to pitch sample PDF. The labelling of the vertical axis, which is unimportant in the present context, was omitted on purpose on this and similar plots to increase the plot size.

What can be seen in Figure 3 is that the F0-values seem to cluster in such a way that an octave (here e. g. the interval between the cluster at 2412 and the cluster at 1209 cents, which spans 1197 cents) is divided into seven intervals of different sizes. This seems to support the interpretation of the clusters as marking the "scale pitches" of the mental tuning template which Artem Erkomaishvili was using during his performance of the chant. Figure 4 shows the melodic line of the top voice of chant no. 2 as a trajectory through the different pitch clusters.

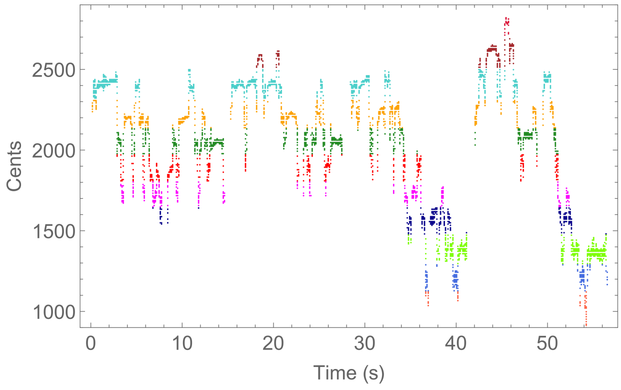


Figure 4. Pitch track of the top voice of chant no. 2, color coded according to cluster membership.

At first glance, one might believe that the spread of the clusters, as indicated by the sample standard deviations in Figure 3, but also the jitter of the pitch track in Figure 4, are rather large since they reach values of one quarter to one half of a semi-tone (25 – 50 cents). However, this is not surprising and must not be seen as a sign of poor pitch control of the singer. For once it is to be expected as an expression of the categorical perception of pitch (e. g. Siegel & Siegel, 1977; Sundberg, 1994). In addition, sliding phases in the beginning of new syllables, breathing, vibrato and consonants all affect the temporal stability of the F0-trajectories. In order to test to what degree these effects, but also the pitch algorithm itself, might affect the determination of the "scale pitches", we performed an alternative pitch determination using the Tony software (Mauch et al., 2014, 2015). In this context, we visually edited the pitch tracks to remove all what could be considered artifacts of sliding phases in the beginning of new syllables, breathing, vibrato, and consonants. Subsequently, pitch tracks as well as notes, yet another way to determine the pitches for this example, were calculated. The resulting histogram is shown in Figure 5.

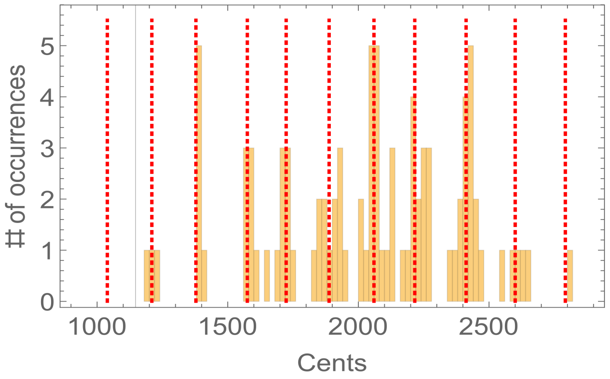


Figure 5. Histogram of note pitches, determined with the Tony software (Mauch et al., 2014, 2015) for the top voice of chant no. 2. The red dotted lines mark the locations of the pitch cluster centers displayed in Figure 1 as determined from the F0-trajectories.

Figure 5 shows that the pitch-cluster mean values determined from the raw F0-trajectories are a reasonable representation of the histogram distribution of the notes, as determined after visual editing of the pitch tracks. As final test of the robustness of the pitch-cluster centers, we performed a k-mean cluster analysis on the individual pre-edited pitch values as determined by the Tony software (based on the PYIN algorithm). The resulting pitch histograms are shown in Figure 6.

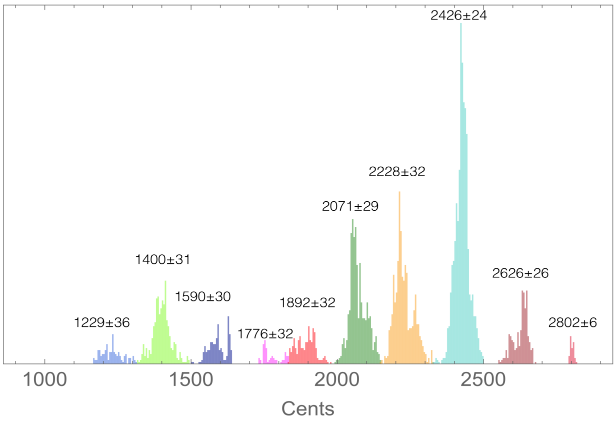


Figure 6. Pitch cluster histograms of the pitch samples in the top voice of chant no. 2 as determined with the PYIN algorithm in the Tony Software (Mauch et al., 2014, 2015. The number on top of each cluster shows the cluster mean and the corresponding cluster standard deviation (in cents).

Overall, comparing the F0-distributions in Fig. 3 and Fig. 6, the set of F0-cluster means in Fig. 3 is shifted by approximately 20 cents towards lower values with respect to the set of cluster means calculated from the pitch samples determined with the Tony software in Fig. 6. This shift might be due to the preprocessing of the pitch trajectories and the removal of presumed artifacts, e. g. glissandi at the beginning of new syllables which tend to start from sometimes rather low pitch values (cf. Figure 4). If one would remove this constant shift, eight of the ten corresponding peaks in the two sets of cluster centers would be less than 10 cent apart, one 15 cent, and one (the smallest cluster in Figure 6 between 1700 and 1800 cents) 33 cent. Based on this result, we might assume that the vast majority of interval sizes, calculated as differences between the cluster centers of neighboring pitch clusters, carry an average uncertainty of less than 10-15 cent.

From the analysis of all chants for which the pitch range of the top voice covers more than an octave (58 out of 101), we obtain 467 intervals between neighboring cluster centers. Their histogram distribution is shown in Figure 7.

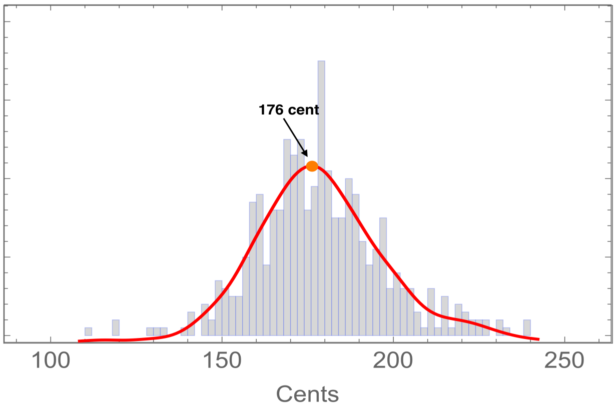


Figure 7. Distribution of possible single-step melodic interval sizes, determined from the top voices of all 58 chants for which for which the pitch range covers more than an octave. The red line corresponds to the 5-point smooth kernel distribution calculated from the histogram data.

No particular preference is seen for any of the interval sizes characterizing seconds in the scale models by Erkvanidze (2002, 2016), by Tsereteli and Veshapidze (2014), or in the tempered diatonic model, respectively. Erkvanidze (2002) suggests that the authentic Georgian tuning system uses interval sizes for single melodic steps that can take values of either 154, 172, or 204 cents. One can see that none of these interval sizes is incompatible with Figure 6, but so are many other intervals between 140 and 220 cents. Tsereteli and Veshapidze (2014) on the other hand suggest a seven-interval scale of equal interval size of 1200/7 = 172 cents. This value, which is also one of the interval sizes in the Erkvanidze model, is actually very close to the peak value of the distribution (176 cents), but there is no visual evidence that Artem Erkvanidze has intentionally tried to achieve this with any precision. Finally, western tempered diatonic scales assume single melodic step sizes of either 100 (semitone) or 200 cents, the first of which is completely absent in Figure 7.

One has to note, however, that Figure 7 does not show the frequency distribution of single melodic steps of notes which were actually sung in a particular chant. It provides an overall view of the possible melodic single-step sizes for the whole corpus. In order to perform a simple test of how different these may be, we used the Tony software (Mauch et al., 2014, 2015) to determine the sung notes (instead of the pitch track centers) in chant no. 2 and to calculate the melodic intervals based on them. The results are shown in Figures 8 and 9.

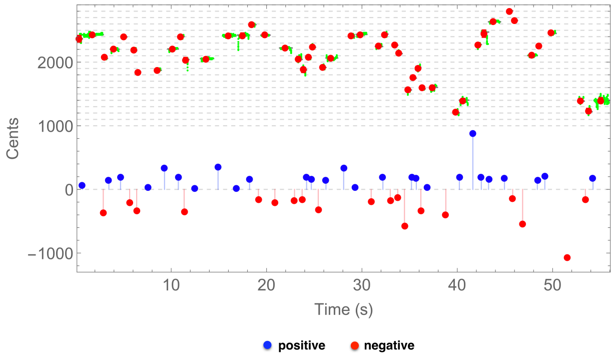


Figure 8. Notes (red blobs plotted at the note pitches in the upper part of the figure, superimposed by the pitch track segments in green) and melodic step sizes (vertical lines in the lower part of the figure). The melodic steps are color coded according to their direction (up- blue, down- red).

In Figure 9a) the distributions of the melodic step sizes is shown independent of direction while in Figure 9b) and c) the distributions are split up according to upwards (b) and downwards (c) movements. As a note on the side, we want to mention that the upwards steps taken by Artem Erkomaishvili in this example seem to be little larger on average than the downward steps. In field observations of traditional village singers in Upper Svaneti (Scherbaum, 2016) this was observed as a systematic feature, which might point to a more general performance element of Georgian vocal music which deserves further study.

It can be seen in Figure 9a) that the central body of the step size distribution for single melodic steps in this example covers a similar range of approximately 140 - 220 cent than the distribution for all possible step sizes shown in Figure 7. In other words, even in a single chant, the variability of melodic seconds is not found to be reduced.

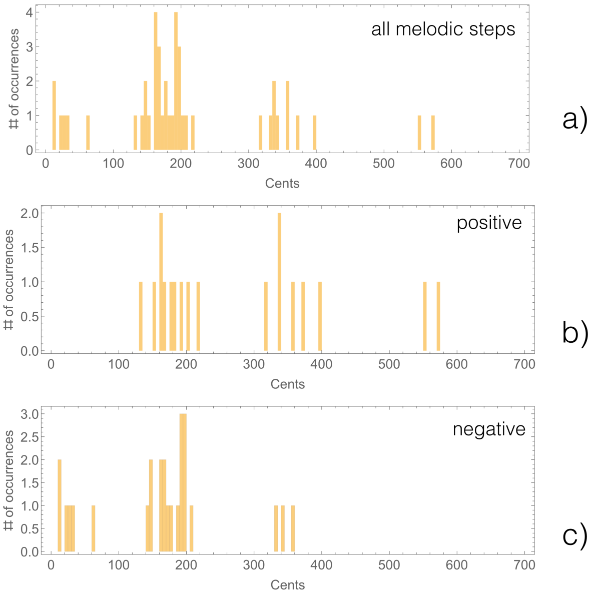


Figure 9. Intervals between notes in the top voice of chant no. 2 (vertical lines) and corresponding histogram. Fig. 9a) shows all steps sizes independent of direction, while these are split up according to positive (b) and negative (c) in the two lower panels.

Figure 7 to 9 suggest that the mental tuning templates Artem Erkomaishvili might have been using during the performance of the chants do not seem to be very rigid regarding the single-step melodic interval sizes, which can also be referred to as “melodic seconds”. For the whole set of 58 chants analysed, the values range approximately between 140 and 240 cent, which corresponds to a semitone in the 12-tone equal tempered scale. There is no visual evidence for any intention to precisely sing any particular melodic scale.

Several questions arise in this context. Does the lack of evidence for precision of the melodic seconds tell us anything regarding the validity of any of the scale models proposed for Georgian music? Is it unintentional or intentional, in other words does it characterize uncertainties or is it actually an important feature of the music and serves a particular purpose? Before we get back to these questions in the discussion section, we are going to look at the other voices and their interaction with each other.

2.2 Harmonic Analysis

To analyze the harmonic tonal organization in the recordings, we realigned the individual voice segments to a common start time. The start and end times for the individual segments were obtained manually and are publiccly available at the website[1] accompanying Müller et al. (2017). First, we aligned all voice segments to a common zero start time. Subsequently, we selected only those F0-samples for which all three voices are active (with valid F0-values). For chant no. 2 this results in the F0-trajectories shown in Figure 10.

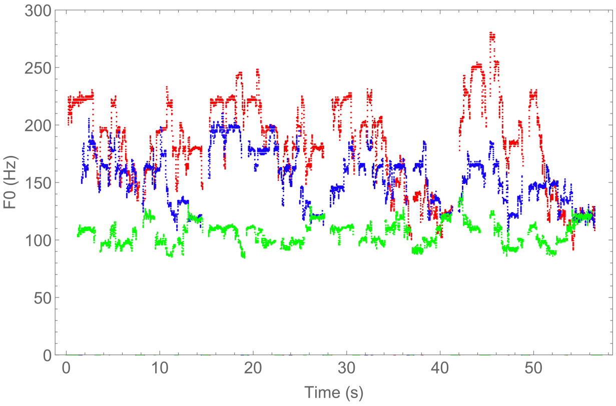


Figure 10. F0 trajectories for the aligned voices of chant no. 2. The top, middle and bass voices are plotted in red, blue, and green, respectively.

Subsequently, we determined the F0-values for all the concomitant pitches from which we calculated the harmonic intervals. These were again subjected to a cluster analysis to quantitatively determine the harmonic structure of the chant (Figure 11).

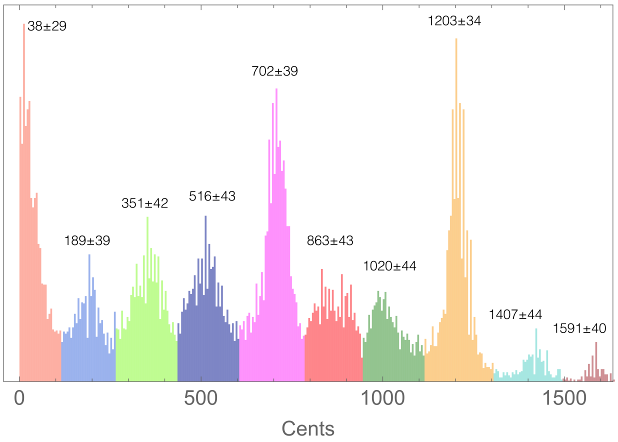


Figure 11. Distribution of concomitant (harmonic) intervals in chant no. 2 and derived clusters thereof. The numbers indicate the cluster means and standard deviations.

One can see in Figure 11 that the most prominent harmonic pitch cluster centers occur around 38 cents, 702 cents (a perfectly justly tuned fifth) and at 1203 cents (a perfectly tuned octave). The harmonic thirds are close to neutral with a cluster center at 351 cents, while the fourths at 516 cents appears sharper than a justly tuned fourth (which would be at 498 cents). Performing the same kind of analysis to all 44394 harmonic intervals in the analysed corpus results in the distribution shown in Figure 12. The general picture remains very similar to Figure 11, except that the harmonic seconds get closer to the tempered value of 200 cents, moving farther away from the distribution of the melodic intervals (cf. Figure 7). Overall, the fifth is the most frequent harmonic interval occuring in the complete corpus.

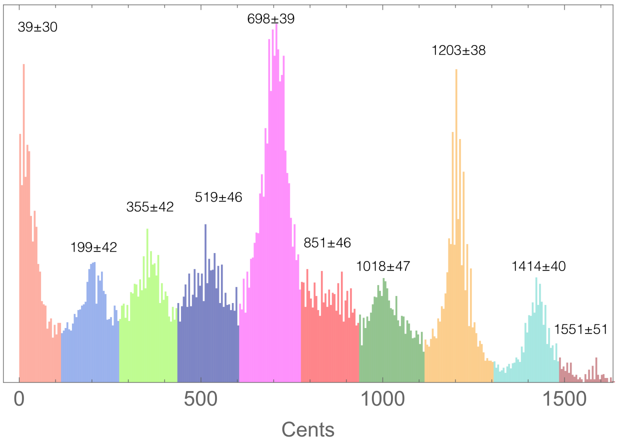


Figure 12. Distribution of all 44394 concomitant (harmonic) intervals in all 58 chants of the corpus for which the top voice covers a range of more than one octave, separated into pitch clusters. The numbers indicate the cluster means and standard deviations.

2.3 Voice Interaction

When Artem Erkomaishvili was singing the middle voice, he was listening to the top voice played back to him from one of the tape recorders. Similarly, he would listen to the recording of the overdubbed top and middle voices when singing the bass. Can one tell from the F0-trajectories, if hearing another voices affects his singing?

[1] https://www.audiolabs-erlangen.de/resources/MIR/2017-GeorgianMusic-Erkomaishvili

If we look at the individual F0-distributions for the different voices shown in Fig. 13, all one can see is that the pitch clusters seem to be pretty much in phase with a similar spread.

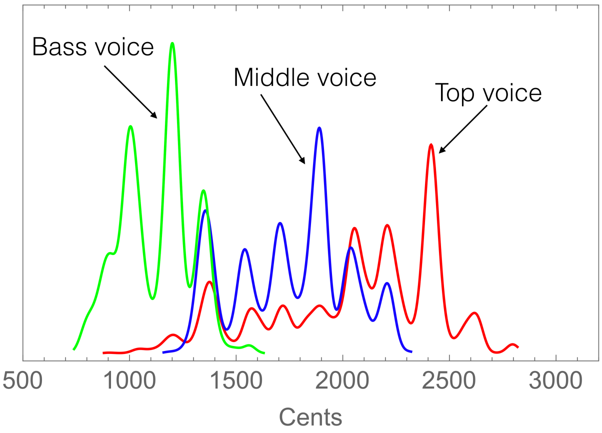


Figure 13. Smooth kernel distributions of the F0 values for the top voice (red), middle voice (blue), and bass voice (green) for chant no. 2.

One way to identify possible voice interactions is by studying the joint distributions of concomitant pitches. These are shown in Figs. 14 to 16 for the middle-top voice, the bass-middle voice, and the bass-top voice pairs, respectively. Each dot represents a pair of simultaneously sung pitches. Jointly sung notes will appear in this plot as a two-dimensional cluster of dots. The x- and y- coordinates of a note cluster should be close to one of the cluster centers for the individual voices shown in Fig. 13. For example the x- coordinates of any of the clusters in Figure 14 (middle against top voice) should be close to one of the peaks of the middle voice (blue curve) in Figure 13, while the corresponding y-coordinates should be close to one of the peaks of the red curve representing the top voice pitches. The reason for this is simply that mathematically speaking the blue and red distributions in Figure 13 are the marginal distributions to the joint distribution of pitch pairs shown in Figure 14. The tilted lines in Figure 14 correspond to different harmonic intervals between the top and the middle voice. The solid black line indicates unisone. So if the two voices are in perfect unisone, the corresponding pitch dot would plot exactly on the solid black line. If the top voice would be exactly 200, 350, 500, or 700 cents above the middle voice, the corresponding pitch dot would plot on the dashed orange, the dashed green, the dashed blue or the solid red line, respectively.

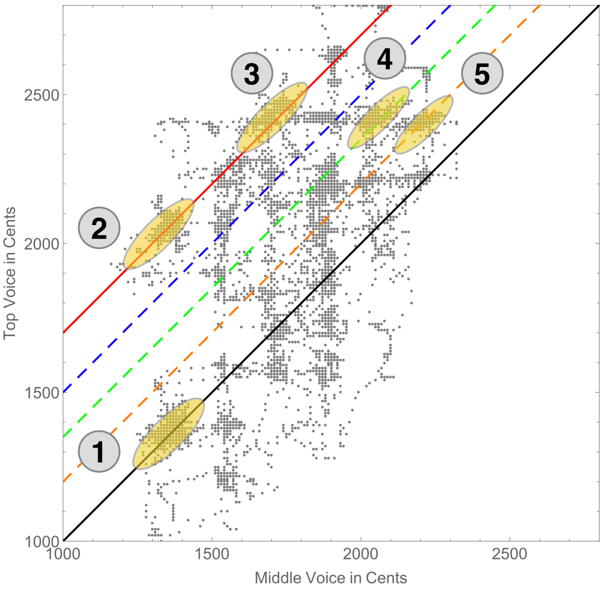


Figure 14. Concomitant middle-top voice pitch pair sample distribution of chant no. 2.

It is the shape of the two-dimensional clusters which tells us if the pitch of the middle voice is influenced by the pitch of the top voice heard. Lets assume, for example, that the top voice sings a note in which the mean pitch is at 1300 cents and fluctuates within a range of ± 20 cents. If the middle voice wants to sing the same note it will also produce a range of pitch values fluctuating by some amount, lets say also ± 20 cents. If the two fluctuations will be complelety independent of each other, say Gaussian, the two-dimensional distribution of pitch pairs will be a two-dimensional Gaussian distribution which would be visible as a distribution around the center point which looks similar in all direction (circular). If, on the other hand, the middle voice would be absolutely stable (no fluctuation at all), one would see a vertical alignment of the two dimensional pitch cluster for that note. If the top voice is stable, but the middle voice fluctuates, then the alignment of the cluster should be horizontal. If, however, the top voice fluctuates by some amount and the middle voice wants to maintain a particular harmonic interval, it must sing in such a way that the middle voice will fluctuate in phase with the top voice by exactly the same amount. In such a case, the note cluster would show an alignment of exactly 45 degrees. In Fig. 14 we can identify several of these structures labeled by numbers. Note cluster 1 in Figure 14, for example, represents a situation in which top and middle voice maintain unisone despite the fact that the voices fluctuate by a considerable amount (by roughly 100 cents). Note clusters 2 and 3 represent situations in which the middle voice sings a stable 5th below the top voice while both voices fluctuate by approximately 100 cents. Note cluster 4 and 5 indicate similar situations for a harmonic neutral third and a harmonic major second, respectively.

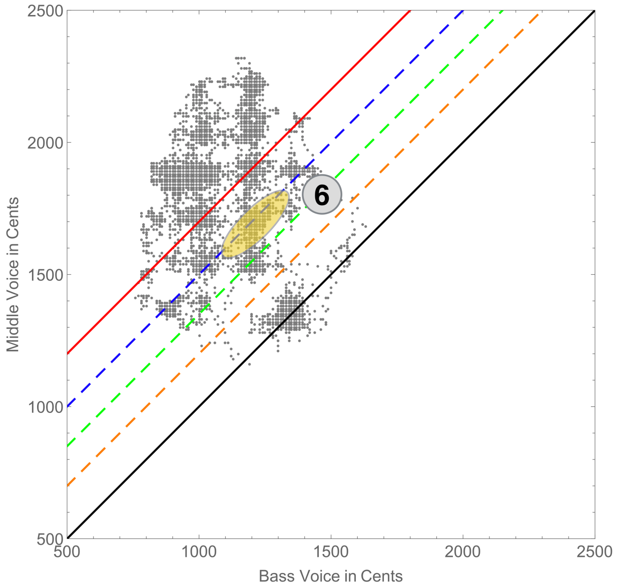


Figure 15. Concomitant bass-middle voice pitch pair sample distribution of chant no. 2.

In the bass-middle voice pitch distribution shown in Figure 15, one can identify more note clusters which are either vertically or horizontally aligned, meaning that there was no or little voice tuning of the bass voice. There is one structure (labeled 6), however, in which a harmonic fourth is attempted to be maintained.

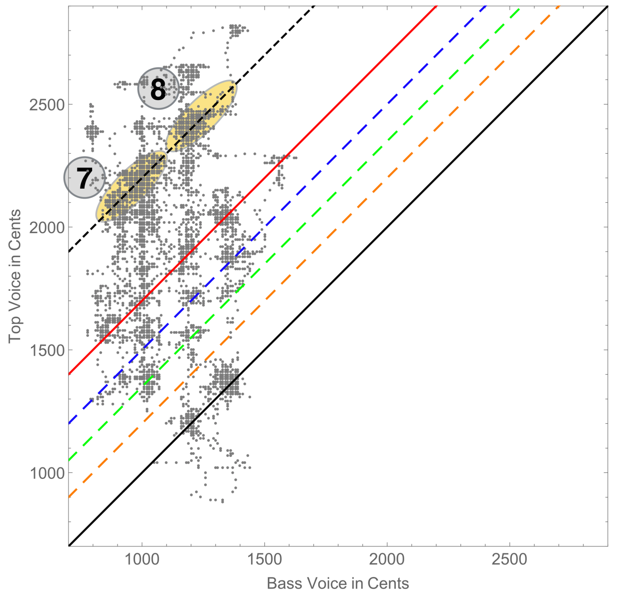


Figure 16. Concomitant bass-top voice pitch pair sample distribution of chant no. 2.

Finally in the bass-top voice distribution, one can see at least two note clusters in which the bass voice tried to maintain an octave to the top voice. Overall, it looks like Artem Erkomaishvili, when singing the bass voice, was switching his attention between the middle and the top voice. When singing the middle voice, on the other hand, he only heard the top voice, so this was his only audible reference which he could relate to. This may explain why Figure 14 shows more note clusters with evidence for voice interactions than Figs. 15 and 16.

3. DISCUSSION AND CONCLUSIONS

With the present study we want to make a contribution to a better understanding of the musical thinking of Artem Erkomaishvili, one of the last master chanters of traditional Georgian chants. Based on a unique set of recordings which was obtained at the Tbilisi State Conservatory in 1966 and for which the F0-trajectories were determined in a prior study by Müller et al. (2017), we investigated the pitch inventories of all three voices of 58 chants separately and jointly. In addition we took a first step at investigating possible signatures of voice interactions between different voices, making use of the special recording setup. In this context it needs to be mentioned, however, that the use of the overdubbing technique, although initiated by Artem Erkomaishvili himself, was new to him (pers. communication by Anzor Erkomaishvili, grandson of Artem Erkomaishvili, 2017). We do not know if this has been influencing the recordings in any way. In any case, there is still much more to be done in the context of trying to understand the influence of voices on each other, in particular on the structural and temporal context in which this happens (cf. Graham, 2013), but this is the objective of a separate study.

Our main results of the analysis of the melodic pitch inventory show that the sizes of melodic seconds vary over a large range from approximately 140 to 240 cents with a peak of the distribution at approximately 180 cents. We do not see a preference for any of the interval sizes characterizing seconds in the scale models by Erkvanidze (2002, 2016), by Tsereteli and Veshapidze (2014), or in the tempered diatonic model, respectively. Loosely speaking, one could characterize the 1966 performance of Artem Erkomaishvili as relaxed regarding the precisons of single melodic steps. In contrast, we observe a high precision when it comes to the harmonic structure of the performance. The harmonic analysis yields a distribution in which precisely justly tuned fifths at 698 cents and octaves at 1203 cents appear as the most frequently intervals. The key to this "melodic flexibility" and "harmonic precision" may lie in the interaction of the voices for which we see clear evidence in the results of the analysis of the joint pitch distributions. There are several cases in which Artem Erkomaishvili maintained particular harmonic intervals despite considerable pitch fluctuations of the individual voices. Therefore, maintaining harmonic precision seems to go hand in hand with the relaxation of melodic precision, which in turn allows for rapid retuning of the voice to maintain an intended harmonic interval.

Relaxing the aim for melodic precision while at the same time aiming at harmonic precision may also relate to the way chants were documented in the past using neumes, which by principle do not allow to document a melody at a very high precision. As discussed in detail in the dissertation of John A. Graham (Graham, 2015), Artem Erkomaishvili used his own neume system, but only for the documentation of the top voice. He is quoted of having told his grandson Anzor Erkomaishvili that "the other voice parts would remember their parts by ear, following the first voice`s lead" (from Graham, 2015). Naturally, if the middle and bass voice are developed by ear from the lead voice, harmonic precision is an asset.

In conclusion, Artem Erkomaishvili's performance in 1966 seems to be characterized by a combination of harmonic and melodic thinking rather than by the single aim for melodic precision.

If this interpretation is correct and if it is valid for traditional Georgian vocal music in general, it would raise the fundamental question whether the concept of a single scale (whatever its parameters are) is appropriate to describe the characteristics of Georgian vocal music. Our results are at odds with any melodic scale model which requires a very high precision in singing the melodic intervals. Since the melodic and the harmonic structure of vocal music does not have to be identical, it seems more appropriate to consider the tonal organization of vocal music as an at least two-dimensional property connecting melodic and harmonic aspects. A very stimulating in-depth discussion of this topic, in particular on the properties of melodic and harmonic seconds, can be found in the paper by Nikolsky (2015).

In a recent paper, Erkvanidze (2016) emphasizes the importance of studying the properties of the old audio recordings of professional master chanters as a means to understand the old Georgians musical system. In his thesis, John Graham writes "any theory must account for both the tuning system heard in the 1966 Erkomaishvili recordings and evidence from earlier singers and other regional chant systems seen in the transcription record." (Graham, 2015). We fully agree with both statements and want to emphasize that in this paper we do not propose any new tuning model. The main aim of the present study is to analyse those acoustical characteristics of the 1966 Erkomaishvili recordings which seem relevant as boundary conditions for model building and provide them for discussion.

Since science usually benefits most from a healthy competition of different ideas and perspectives, we also invite other researchers to test their models (or develop new ones) using the F0-trajectories of the Erkomaishvili recordings, which for this purpose have been made publicly available[1].

4. REFERENCES

Erkvanidze, M. (2002). On Georgian Scale System. In The First International Symposium on Traditional Polyphony: 2-8 September, 2002, Tbilisi, Georgia (pp. 178–185).

Erkvanidze, M. (2016). The Georgian Musical System. In 6th International Workshop on Folk Music Analysis, Dublin, 15-17 June, 2016. (pp. 74–79).

Gelzer, S. (2002). Testing a scale theory for Georgian folk music. In The First International Symposium on Traditional Polyphony: 2-8 September, 2002, Tbilisi, Georgia (pp. 194–200).

Gogotishvili, V. (2004). On authentic and plagal types of monotonic (non-octave) scales in Georgian traditional vocal polyphony. In Proc. of the Second International Symposium on Traditional Polyphony, 23-27 September, 2004, Tbilisi, Georgia (pp. 218–226).

Graham, J. A. (2013). Unity and Variety in Orthodox Music: Theory and Practice. In I. M. and M. Takala-Roszczenko (Ed.), *Proceedings of the Fourth International Conference on Orthodox Church Music University of Eastern Finland*. Joensuu, Finland: The International Society for Orthodox Church Music (ISOCM).

Graham. J. A. (2015). The transcription and transmission of Georgian Liturgical chant. PhD thesis, Princeton University.

Jgharkava, I. (2016). Pearls of Georgian Chant, CD produced by the Georgian Chanting Foundations & Tbilisi State Conservatoire.

Kawai, N., Morimoto, M., Honda, M., Onodera, E., & Oohashi, T. (2010). Study on sound structure of Georgian traditional polyphony. Analysis of its temperament structure. In The Fifth International Symposium on Traditional Polyphony, 4-8 October, 2010, Tbilisi, Georgia (Vol. 1, pp. 532–537).

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

Müller, M. (2015). Fundamentals of Music Processing — Audio, Analysis, Algorithms, Applications. Springer Verlag, ISBN: 978-3-319-21944-8, 2015

Müller, M., Rosenzweig, S., Driedger, J., & Scherbaum, F. (2017). Interactive Fundamental Frequency Estimation with Applications to Ethnomusicological Research. Submitted to *Conference on Semantic Audio, 2017 June 22 – 24, Erlangen, Germany* (8 pages).

Nikolsky, A. (2015). Evolution of tonal organization in music mirrors symbolic representation of perceptual reality. Part-1: Prehistoric. *Frontiers in Psychology*, *6*(OCT), 1–36. http://doi.org/10.3389/fpsyg.2015.01405

Scherbaum, F. (2016). On the benefit of larynx-microphone field recordings for the documentation and analysis of polyphonic vocal music. Proc. of the 6th International Workshop Folk Music Analysis,15 - 17 June, Dublin/Ireland, 80–87.

Shugliashvili, D. (2014). Georgian Church Hymns, Shemokmedi School (pp. XXIII–XXIX).

Siegel, J. A., & Siegel, W. (1977). Categorical perception of tonal intervals: Musicians can't tell sharp from flat. *Perception & Psychophysics*, *21*(5), 399–407.

Sundberg, J. (1994). Perceptual aspects of singing. *Journal of Voice*, *8*(2), 106–122.

Tsereteli, Z., & Veshapidze, L. (2014). On the Georgian traditional scale. In The Seventh International Symposium on Traditional Polyphony: 22-26 September, 2014, Tbilisi, Georgia (pp. 288–295).

Westman, J. (2002). On the problem of the tonality in Georgian polyphonic songs: The variability of pitch, intervals and timbre. In The First International Symposium on Traditional Polyphony: 2-8 September, 2002, Tbilisi, Georgia (pp. 212–220).

[1] https://www.audiolabs-erlangen.de/resources/MIR/2017-GeorgianMusic-Erkomaishvili

INTERVAL STRUCTURES IN GNAWA MUSIC: DYNAMICS OF CHANGE AND IDENTITY

Lorenzo Vanelli

University of Bologna
lorenzo.vanelli3@unibo.it

ABSTRACT

This article's main aim is to outline the intervals between the pitch of the sounds that compose the modal structure that is the most commonly used by the *Gnawa* of Morocco. Every Moroccan brotherhood has its music, and the intrinsic details of those traditions are inextricably tied to the history of each of them. To analyze the details of the musical structures and techniques used in the brotherhood's rites contributes to the comprehension of their present and could give insights on their past.

We will discuss the interval structure of the modal scale through the techniques that the *Mâallem* implements when playing the *guembrì* and through the analysis of the audio recordings, with a particular focus on the melodic lines of the singers.

The results obtained could contribute to the confirmation of the historical and contemporary distinctiveness of the *Gnawa* from the other Moroccan brotherhoods. The data collected in the analysis could also make way for a discussion about the transformations and the unbalanced negotiation occurred in the process of commercialization of the *Gnawa* music across the global market.

1. INTRODUCTION

In Morocco there are a number of *sufi* brotherhoods (*ṭarīqa*), whose ritual activities are strongly intertwined in the daily life of the inhabitants. The brotherhoods differentiate themselves from one another on the base of ceremonial elements, dress codes, territorial localization, sanctuaries (*zaouïa*), and musical practices. In fact, the *Gnawa* brotherhood tends to be significantly more distinct than others, especially in regard to music.

The historical reasons of this diversity have been investigated, and recent research (including oral narratives) show that it could be related to the ethnic and social provenance of the original nucleus of founders of the brotherhood. Many scholars[1] have found evidence supporting that the *Gnawa* tradition emerged from the constitution of a Moroccan-styled brotherhood at the hand of the black ex-slaves deported to Morocco from the Western and Central African countries. The history of the *Gnawa* brotherhood is that of the institutionalization of the presence of those minorities in the fabric of Moroccan society through the construction of a religious structure that is the result of a long and complex negotiation with local practices. The specificities of the music by the *Gnawa* brotherhood offer a perspective as to how their history of slavery has become incorporated and reclaimed into their identity, marking them from other *ṭarīqa* that do not share the same past. Their music represents resistance and expression of cultural affirmation against historical and contemporary racism that they have faced and are facing.[2] It talks about their resistance to historical adverse situations where their culture risked to be erased, but survived to reaffirm itself in the construction of the brotherhood. Today the *Gnawa* brotherhood is seen as one of the most important of Morocco, to the extent of being an example for other brotherhoods of Moroccans.[3] They were able to build up a common cultural heritage that celebrates their empowerment and survival.

For these reasons, a proper study of the specific elements that compose the *Gnawa* culture adds knowledge not only to the description of the contemporary situation, but also to the understanding of the historical process that led to the cultural and social relations that we see today.

2. CONTEXT OF RESEARCH

The *Gnawa* music tradition distinguishes itself, apart from the characteristic use of specific instruments and rhythms, also through the use of particular modal scales whose interval structure is substantially extraneous to the Moroccan context.

The discussion presented in this article is mainly based on the data and materials that I collected during two research trips for a total of five months in 2016 and 2017 to Essaouira, Marrakech, Casablanca and Rabat.[4] The princi-

[1] For more informations on the history, the rituals and the culture of the *Gnawa*, and for more details on the variety of provenances of the founders of the brotherhood, see Becker (2011), Bentahar (2010), Chlyeh (1998), El Hamel (2013), Pâques (1991), Shaefer (2015), Sum (2011, 2013), and Turchetti (2015).

[2] Even if the recent political activity in Morocco is going in the direction of properly confronting racism and discrimination in the country, the situation is still dire. For more information on the subject, I suggest to visit the GADEM website, and De Haas (2014), El Hamel (2002, 2013), Marouan (2016), and Timéra (2011). Additional information on the subject can also be found in the articles cited in the previous footnote.

[3] See El Hamel (2008: 249 et seq.).

[4] The research trip was coordinated by Prof. Staiti of the University of Bologna as part of the European project D.R.U.M. I also owe to prof. Staiti the merit of having chosen the theme of the research and outlined

ISBN: 978-84-697-2303-6

pal informers that I worked with have been Khalid Amenhouce (47 years old, *Mâallem* based in Essaouira, apprentice of the recently deceased Mahmoud Guinia), Yassine Boubker (21 years old, he plays professionally the *guembrì* since the age of 16), Azouz Soudani (57 years old, affirmed professional player with a long experience), and Mahmoud (55 years old, *guembrì* constructor based in Essaouira's medina). To verify the data collected against a broader documentation sample, I also analyzed a vast number of recordings of *Gnawa* performances that I bought directly in Morocco[5] or found on the internet.[6]

3. INSTRUMENTS, TECHNIQUES, DATA

The *guembrì* is a lute with three strings, called *dir*, *tehtia* and *westia*.[7] The *dir* is the lowest string in terms of pitch: *tehtia* and *westia* are tuned respectively a fourth and an octave higher. The *westia* is shorter than the other two, stands in the middle of them, and is attached to the bridge right next to the third finger position on the *tehtia*. The main modal structure used in *Gnawa* music (although not the only one) requires the musician to obtain at least seven notes from the instrument: *dir*, *tehtia* and *westia* open string, plus one finger position on the *dir* and three on the *tehtia*. I measured the length of the strings on all the instruments used by the musicians I talked with, and asked them to show me the finger positions that they reach for on the strings to obtain the notes, then measured that distance from the bridge. To cope with eventual differences between the base length of the strings (as the instruments are handmade the strings could differ in length even a few centimeters) I calculated the mean values of the strings base lengths, then modified the data of the finger position accordingly through a simple proportion. In the end I calculated again the mean values of those positions between the different musicians. I was not particularly surprised to find that the four professional musicians I interviewed showed me exactly the same proportions between the finger positions. In the Tables n. 1 and 2 I show the mean values of the length of the vibrating part of the string for each note. I calculated the distance expressed in cent between *dir*, *tehtia* and *westia* verifying the tuning of the strings, whereas I calculated the distance in cent of the note obtained with each finger position on a string from the fundamental frequency of the open string itself through the formula:

$$f = \frac{1}{2l}\sqrt{\frac{T}{\mu}} \qquad (1)$$

Where f is the frequency of the sound produced, l the length of the vibrating part of the string, T is the tension and μ is a constant tied to the material that composes the string. As the last two can be considered constant on the same string, by calling k the result of the root of their ratio we could describe the frequency of a note coming from a given finger position (or of the open string itself) as

$$f_1 = \frac{1}{2l_1}k \quad \text{and} \quad f_2 = \frac{1}{2l_2}k \qquad (2)$$

Then we can recur to this formula to calculate the distance in cents between two frequencies:

$$\text{Dist} = 1200\log_2\left(\frac{f_2}{f_1}\right) \qquad (3)$$

By substituting the two formulas for f_2 and f_1 from the (2) in the (3), and making the appropriate simplifications, we obtain the final formula that I used:

$$\text{Dist} = 1200\log_2\left(\frac{l_1}{l_2}\right) \qquad (4)$$

Position	*Length [cm]*	*Absolute error L [cm]*	*Interval from modal center [cent]*	*Absolute error on the interval [cent]*	*Interval*
Dir open string	*83*	*0,5*	*0*	*0*	*Tonal center*
Dir first position	*75,5*	*0,5*	*163,9*	*2,1*	*Raised neutral second*
Tehtia open string	*83*	*0,5*	*500*	*0*	*Perfect fourth*
Tehtia first position	*74*	*0,5*	*698,7*	*2,5*	*Perfect fifth*
Tehtia second position	*65*	*0,5*	*923,2*	*5,8*	*Raised major sixth*
Tehtia third position	*62*	*0,5*	*1005,0*	*7,1*	*Minor seventh*
Westia open string	*62*	*0,5*	*1200*	*0*	*Perfect eight*

Table 1. Length of vibrating part of strings on the *guembrì*

the hypothesis of which this brief relation is meant to be both a development and a test. Although in this article I will generalize by referring the modal structure to the broader Moroccan *Gnawa* heritage, it is important to remember that there are conspicuous regional variations and that the cultural heritage described here is more commonly found in the northwestern area of the country.

[5] Many musicians promote themselves through selling homemade recordings on cds or directly by handing out mp3 files through pen drives.

[6] Similarly to the homemade production of cds and mp3, musicians use the most common internet service providers (namely Youtube and Facebook) to demonstrate their capabilities and promote their music.

[7] Sum (2012: 128) calls them *zir*, *tahtiya* and *ndui*, saying that those word mean "fater", "mother" and "child" respectively. My findings differ from the author's: see the appendix for more information.

Another opportunity to verify the results shown in Table 1 came from the analysis of the finger positions used by *guembrì* players on another model of instrument called *hajhuj*. This one is constructed in the same way than the *guembrì*, but is only used in particular ritual contexts.[8] It is also a bit smaller, and has a different system of connection of the strings to the top of the bridge. Normally *dir* and *tehtia* have different lengths, but on the *guembrì* the difference in length is annulled by the fact that the lace blocking the second also holds down the first, making the two open string lengths just equal. This does not happen in the smaller *hajhuj*, where the strings are normally directly entwined in locking mechanisms. That leaves the lengths of the open strings unmodified by laces, thus making them differ. In Table 2 we can see that, even with this difference, and keeping into account the absolute error tied to the derived measure in cents, the musicians accommodate for the difference by varying the finger positions accordingly, obtaining positions that still reproduce the same intervals seen in Table 1.

Position	*Length [cm]*	*Absolute error on L [cm]*	*Interval from the modal centre [cent]*	*Absolute error on the interval [cent]*	*Interval*
Dir open string	76	0,5	0	0	*Modal centre*
Dir first position	69	0,5	167,3	2,3	*Raised neutral second*
Tehtia open string	74	0,5	500	0	*Perfect fourth*
Tehtia first position	66	0,5	698,0	2,8	*Perfect fifth*
Tehtia second position	58	0,5	921,8	6,5	*Raised major sixth*
Tehtia third position	55,5	0,5	998,0	7,9	*Minor seventh*
Westia open string	52	0,5	1200	0	*Perfect eigth*

Table 2. Length of vibrating part of strings on the *hajhuj*

4. VOICE SPECTROGRAMS

We can also confirm the intervals identified studying the techniques of use of the *guembrì* and *hajhuj* through another way: by the relations between the frequencies reached by the singers, evaluated from the spectrograms of the recordings.[9] For an easier reading of the Figures, I slightly modified the image output by adding a horizontal grid[10] and by highlighting in green the traces of the melodic movement, carefully confronting aural and visual information.

In Figure 1 we see the spectrogram of a melodic passage from a recording of *Šalaba*, one of the songs where the modal scale discussed here is used.[11] In the Figure we see the passage from the first hemistich of a verse, concluding in the center after the long note, to the second hemistich. The grid added on top of the image signals the semitones of a tempered scale, and the red lines show the modal center and the perfect fourth, fifth and octave. We can observe the singer going for a precise intonation of the perfect octave, minor seventh, perfect fifth and fourth, but also the particular tuning of the second from the modal center, in the lower right corner, as a slightly raised neutral second. Note the slight upward movement of the voice, that points even more in the direction of the raised neutral tuning. In this module we cannot directly see the sixth from the modal center, but we have a visibly raised major ninth in the first arch. The fact that a raised major ninth and a raised neutral second share the same space is understandable if we keep in mind that those songs tend to have a movable modal center that the singer relocates in relation to the flow of the verses and hemistiches. In this case, we could interpret the first (left) part of the verse as having a modal center positioned at a perfect fourth from the final note of the module. In other words, the modal center of the first hemistich is on the note of the open *tehtia* string, while the modal center of the second one is on the open *dir* string, which makes even more sense as it relates the discourse to local instrumental practice. Considering the pitch of the *tehtia* as the temporary modal center of the first part of the verse, then the pitch of the note achieved in the first arch is exactly at an interval of raised major sixth from it. Therefore, Figure 1 confirms the description of the interval structure of the modal scale given on the base of the prin-

[8] More research should be done on the subject, as the use of the *hajhuj* is more elusive, but it seems that it is only used in rites focused around female *jinns*. I thank Silvia Bruni for the information on the subject.

[9] I obtained the spectrograms from Sonic Visualiser, a software produced by the Queen Mary University of London: for more information, see Cannam (2010).

[10] The software produces png images without the grid, so with the help of Alberto Malagoli, an engineer from the University of Modena, I created a simple html-based program that adds a modifiable grid on top of the png file. The program can be found on the website http://alpert.altervista.org/LorenzosTesi/

[11] See appendix for more information on the song. I chose this one because in its first part it is accompanied only by a sparse rhythmical clapping of hands: the lack of *qraqab* makes the spectrogram clearer.

cipal finger positions on the *guembrì*, previously described, but it also suggests interesting details about the relation between the coordination of moving modal centers, the lyrics content and structure, and the instrumental techniques.

In Figures 2 and 3, instead, we can see the same module of the same song interpreted by *Mâallem* Hamid el Kasri, in one of his recordings[12] for the circuit of the world music production. In this recording, other than the traditional instruments, a keyboard tuned on the tempered scale complements the accompaniment. As we can clearly see, the singer adapted the pitch of the notes to the tempered scale suggested by the keyboard, thus flattening the specificities of the *Gnawa* modal scale. In particular, he adapted the interval of a raised neutral second from the final note to a major second (Figure 2, lower right corner). The interval of a raised major ninth (Figure 3, upper left corner) has been accommodated in the same way, resulting in a modal structure of the piece flattened to that of a dorian mode.

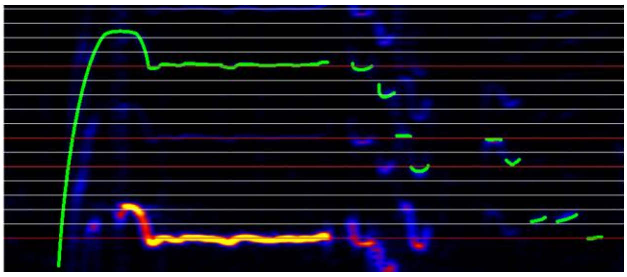


Figure 1. Spectrogram of the melodic movement of the voice in one of the *Šalaba* modules, recording by *Mâallem* Khalid Amenhouce

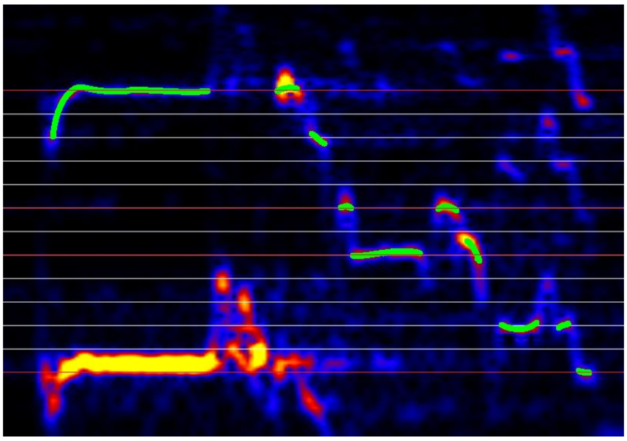


Figure 2. Spectrogram of the melodic movement of the voice in one of the *Šalaba* modules, recording by *Mâallem* Hamid el Kasri

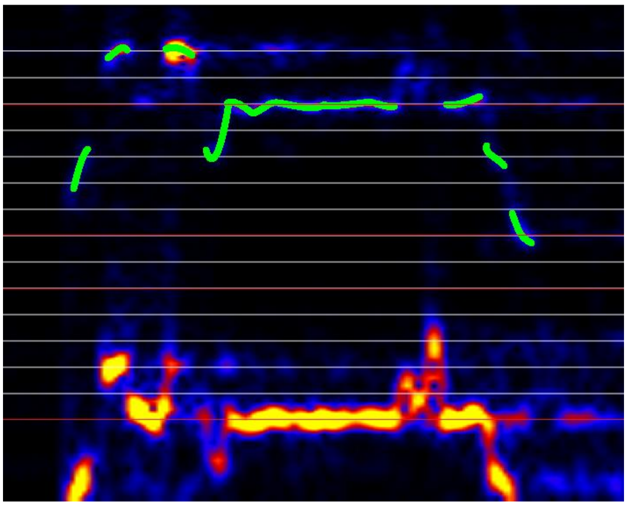


Figure 3. Spectrogram of the melodic movement of the voice in one of the *Šalaba* modules, recording by *Mâallem* Hamid el Kasri

5. MODAL STRUCTURES

From what we've gathered, *Gnawa* music in traditional contexts is mainly based on a modal structure only partially equivalent to a dorian mode without the third degree, where the second and sixth degree are toned in a complex way that is not in line with that built on a tempered scale.

The only researcher who described this modal structure and its variations is Maisie Sum, who published the results of her analysis in two articles (2011: 88; 2013: 157) that partially resume the informations more thoroughly explained in her doctoral dissertation (2012: 128-136).

Sum's articles do not delve into the discussion of the intervals as they are limited to a general description of the pentatonic scale. Her dissertation, on the other hand, presents a detailed study of the interval structure, which shows substantial similarities to the data that I collected. Nevertheless, the author's choice to describe the pitches of all the positions in relation to the closest tempered notes unnecessarily complicates the description of the actual intervals between them: for this reason, I preferred to discuss the intervals simply by taking into account their relation to the fundamental pitch. Moreover, the author claims that the *westia* "is always played open" (2012: 128), whereas at least two additional positions, and the relative high-pitched notes, are sometimes obtained on it. The musicians that I interviewed could not show them to me one by one, as they used them only in the climaxing moments of the performance and during very fast melodic movements. For these reasons the Tables that I presented did not cover those positions. Through the analysis of the recordings, however, those notes seem to be tuned at an interval of a fourth and a second – variably interpreted – from the *westia* note, thus replicating the mode.

[12] The recording is the seventh track contained in the cd Hamid el Kasri & Issam-Issam, *Saha Koyo: Jazz Gnawa*, published by the Societè Artistique Et Culturelle Audio Visuelle, SACAV S 127 (2006)

The interval structure described here, like the ones in Sum (2012: 127 et seq.), are constructed in relation to the pitch of the *dir* string, which acts as the main fundamental of the mode: as it is the lowest note that a *guembri* could produce, musicians cannot go below it. In other songs, whose use is less widespread, musicians instead use a modal structure that constantly refers to the *tehtia* as the main fundamental, thus providing the option to go below it using the *dir*. From what I was able to observe those modes substantially reconstruct the same interval structure discussed, but transposed up a fourth. Musicians thus play on the *dir* notes that have the same structural function of the first and second position on the *tehtia* in the mode discussed in the Tables: a fourth (open string *dir*) and a major second (first *dir* position more advanced on the bridge) below the *tehtia* pitch. The musicians also move the first position on the *tehtia* slightly back to accommodate the pitch of the neutral second. Some songs of the *Gnawa* repertoire that use this mode also add a position on the *tehtia* that produces a major third from the fundamental, and sometimes the first position on it gets lowered even more until it's somewhat close to a minor second from the fundamental. The relative rarity of those songs and the time constraints of my research did not permit me to collect enough documentation to be able to give a proper account of those modes.

6. CONCLUSIONS

This paper has briefly assessed how musicians construct the intervals that compose the main modal structure used in *Gnawa* music in a strikingly recognizable way. That structure functions as an identity marker for those who affiliate with the *Gnawa* brotherhood. They recognize it and reproduce it through their music, alongside all the other elements that compose their rituals, not only to achieve the performance and the intervening people's needs, but also to reproduce and regain possession of the heritage of the founders of the brotherhood in a culturally accepted way.

As briefly discussed in the introduction, this discourse is especially relevant in the contemporary situation. Notably, not all *Gnawa* musicians (and not all people who attend to *Gnawa* rituals) are from or can prove with certainty to have ancestors from the Western or Central African countries, but strikingly all of the people I interviewed strived to demonstrate to me that some branch of their family had some kind of ties to those regions. Even if they do not or cannot prove those ties, by affiliating with a brotherhood whose name etymologically means "black people"[13] they are still identifying themselves in a particular manner, that further enters in a complex dialogue with the Moroccan social context. *Gnawa* musicians have earned the favor of local and international tourism, and despite racial discrimination, they achieved fame, acceptance, and in many cases a strong sense of affiliation from Moroccans: and that happened also because of the specific qualities of their music.

What is also happening in recent years, though, is that European and North American producers have been appropriating the local *Gnawa* expertise and culture, while some local music producers are struggling to accommodate to the aesthetics of the international audience, without any considerable monetary return for the participants. A simple but self-explanatory example of a widespread behaviour: a 17 year old black man I met in Essaouira, named Abdou, son of a Berber family who has lived for nearly three generations in Morocco after moving from Sudan and who now resides on the Atlas mountains, is an extremely skilled *Gnawa* dancer and *qraqab* player. During the 2016 International Festival of *Gnawa* Music in Essaouira he danced, played and sung for four days and four nights. Because of his expertise, most of the times he performed as soloist in front of the dancers group, led the parades, and animated with his infinite energy the performances. The producers of the festivals promised to pay him five hundred dirhams, or about fifty euros. Three months after the festival he was still waiting for the payment.

European and American music industries abuse the musicians by poorly compensating them for their expertise while pretending to give them international visibility. The music managers treat the musicians without regard for the complexities of their traditions, and this fetishization probably operates as a further justification for the musicians' low compensation. When participating in those festivals, *Gnawa* musicians tune down, flatten and abandon the same differences and elements that distinguish them, while at the same time they exaggerate other elements for the sake of spectacularization. As seen in the analysis of Hamid's recording, all the nuances of the *Gnawa* modal structure were completely annulled and absorbed in the tempered scale, leaving only a short-lived and culturally opaque musical result.[14]

Even if it is not possible to know the future impact of those festivals on the local practices of the *Gnawa* brotherhood, today the *ṭarīqa* is still renowned and beloved by Moroccans, who seek the *Mâallem*'s expertise for their private *lila*, which still produce the major income for those musicians. In the *lila*, and in the less formal and secular performance opportunities for tourists in locally-managed restaurants, the *Gnawa* are still engaging on their terms for social and financial gain, while keeping their tradition alive.

The interval details discussed in this article, other than pointing out some transformations occurring inside some of the contexts where *Gnawa* musicians operate, open up at least two observations. This data could be a useful documentation in researches about the positionality of *Gnawa* musicians in Morocco, and about the geographical prove-

[13] See the complete etymology in El Hamel (2008)

[14] For a general discussion of the system of music festivals in the field of ethomusicology, see Staiti (2013).

nance of the founders of the brotherhood. The ethnomusicological attention to the technical details of the construction of the *Gnawa* repertoire could also make way for a deeper and clearer research that interrogates the relations between the elements through which the rites are performed and the technical nuances that construct the musical performance.

7. APPENDIX: DARIJA TERMS

A foreword: *Darija* is an oral language, hence the transcriptions of the words may vary depending on the nationality of the transcriber. Here I try to offer the reader a concise explanation of the words used in the article, while trying to follow the most commonly accepted way of writing them down.

Mâallem: Master. In the ritual contexts he covers the roles of main singer, *guembrì* player and knowledgeable controller of the flow of the rite. While the vast majority of the *Gnawa* musicians are men, women are not forbidden to become *qraqeb* or *guembrì* players and singers, and at the end of a long training, *mâallem*.

Guembrì: a drum-lute with three strings mounted on a neck embedded in a hollow half log covered with a stretched camel skin. The bridge transmits the vibration of the strings to the camel skin, which the player can also beat with his fingertips at the end of the strumming movement. The skin has a hole at the bottom for acoustic reasons. The neck does not go through the whole log but stop short a few centimeters from the lower part, right below the hole in the skin: the lower ends of the strings are tied here. The three strings have different length (one is much shorter than the others), and the pitch of the instrument is overall rather low. During the evocation of the spirits the *Gnawa* players attach a metallic palette at the head of the instrument with small rings embedded in it, to add some metallic timbre to the instrument. A smaller version of the *guembrì* used only in specific ritual contexts is called *hajhuj*.

Qraquab: set of two metallic castanets, with two hollow cavities each. *Gnawa* people say that they reproduce the sound of the chains that tied the hands and feet of the slaves.

Dir, tehtia, westia: the names of the strings of a *guembrì*, ordered by pitch from lowest to highest. The names and meanings that I offer here differ from those described in Sum (2012: 217), both in the way I spell the phoneme and in the meaning. When the *Mâallem* is holding the instrument the position of the strings, seen from the perspective of a frontal onlooker, is as follows:

_______________ *dir*
_________ *westia*
_______________*tehtia*

The information is particularly relevant because the names relate to their position. *Tehtia* literally means "something below", while *westia* means "something in the middle". As for *dir*, it means "the space between": it could be between two mountains, a hollow or a valley. As the *dir* has the deepest sound (lowest pitch) of the three, the topographical reference makes sense, in a metaphorical way. I thank Badr Dammouch and Zakaria Rhani for their linguistic commentary on the meaning of those terms.

Ṭarīqa: a *sufi* brotherhood.

Zaouïa: a house-sanctuary owned by the descendants of the founder of a *Ṭarīqa*, or the construction where his body was entombed. The *zaouïa* is the place where the major indoor rituals are performed, although they are also carried out inside private houses.

Lila: "night". It's the term used to designate the rituals, as they span through the night hours from evening to morning.

Šalaba: a song performed in the first part of the *lila*, before the summoning of the spirits. It relates to the ritual pouring of mint tea for the guests. The lyrics of the song invite everyone to participate by drinking the sacred beverage: "Šalaba titara difu li Allah". "Šalaba" is the *Gnawa* term for "to drink"; "titara" recalls the sound of the bells of the water sellers in Moroccan medinas. Among the *Gnawa*, "titara" could also be directly translated to "tea", seen not as a normal drink, but as one charged with ritual meaning, so the whole sentence could be translated as "Drink the ritual and celestial tea, oh guests of Allah". I thank Silvia Bruni for her research and insights regarding the ritual use of the song, the lyrics and their meanings.

8. REFERENCES

Becker, C. (2011). Hunters, Sufis, soldiers and minstrels: The diaspora aesthetics of the Moroccan Gnawa". *Anthropology and Aesthetics*, 59/60: 124-144

Bentahar, Z. (2010). The visibility of African identity in Moroccan music: From Gnawa to Ghiwane and back. *Wasafiri*, 25(1): 41-48

Cannam, C., Landone, C., Sandler, M. (2010). Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files. *Proceedings of the ACM Multimedia 2010 International Conference*.

Chlyeh, A. (1998). Les Gnaoua du Maroc: Itinéraires initiatiques, Transe et Possession. Grenoble: Éditions La Pensée Sauvage,

De Haas, H. (2014). Morocco, setting the stage for becoming a migration transition country? Research published on 19/03/2014 on the MPI page http://www.migrationpolicy.org/article/morocco-setting-stage-becoming-migration-transition-country

El Hamel, C. (2013). Black Morocco: a history of slavery, race and Islam. New York: Cambridge University Press.

El Hamel, C. (2008). Constructing a diasporic identity: tracing the origins of the Gnawa spiritual group in Morocco. *The Journal of African History*. 49(2): 241-260

El Hamel, C. (2002). 'Race', slavery and Islam in Maghribi Mediterranean thought: the question of the Haratin in Morocco. *The Journal of North African Studies*, 7(3): 29-52

Marouan, M. (2016). Incomplete forgetting: Race and slavery in Morocco. *Islamic Africa*. 7: 267-271

Pâques, V. (1991). La religion des esclaves, recherche sur la confrérie marocaine des Gnawa. Bergamo: Moretti e Vitali Editori.

Shaefer, J. P. R. (2015). Observations on Gnawa healing in Morocco: Music, bodies and the circuit of capital. *Performing Islam*. 4(2): 173-182

Staiti, D. (2013). Interculturalità? I concerti, la world music e l'etnomusicologia. *L'etnomusicologia italiana a sessanta anni dalla nascita del CNSMP (1948.2008)*, Roma: Accademia Nazionale di Santa Cecilia.

Sum, M. (2011). Staging the sacred: musical structure and processes of the Gnawa lila in Morocco. *Ethnomusicology*. 55(1): 77-111

Sum, M. (2012). *Music of the Gnawa of Morocco: evolving spaces and times*, Electronic Theses and Dissertations, University of British Columbia.

Sum, M. (2013). Music for the Unseen: Interaction between Two Realms During a Gnawa Lila. *African Music: Journal of the International Library of African Music* 9(3) pp. 151–182

Timéra, M. (2011). La religion en partage, la <couleur> et l'origine comme frontière: Les mingrants sénégalais au Maroc. *Cahiers d'études africaines*. 51(201): 145-167

Turchetti, A. (2015). Un rituale sincretico e polisemico: la lila degli Gnawa marocchini. *Anthrocom Online Journal of Anthropology*. 11(2): 127-154

RHYTHMIC PATTERNS IN RAGTIME AND JAZZ

Daphne Odekerken
Utrecht University
d.odekerken@uu.nl

Anja Volk
Utrecht University
a.volk@uu.nl

Hendrik Vincent Koops
Utrecht University
h.v.koops@uu.nl

ABSTRACT

This paper presents a corpus-based study on rhythmic patterns in ragtime and jazz. Ragtime and jazz are related genres, but there are open questions on what specifies the two genres. Earlier studies revealed that variations of a particular syncopation pattern, referred to as 121, are among the most frequently used patterns in ragtime music. Literature in musicology states that another pattern, clave, is often heard in jazz, particularly in songs composed before 1945. Using computational tools, this paper tests three hypotheses on the occurrence of 121 syncopation and clave patterns in ragtime and jazz. For this purpose, we introduce a new data set of 252 jazz MIDI files with annotated melody and metadata. We also use the RAG-collection, which consists of around 11000 ragtime MIDI files and metadata. Our analysis shows that syncopation patterns are significantly more frequent in the melody of ragtime pieces than in jazz. Clave on the other hand is found significantly more in jazz melodies than in ragtime. Our findings show that the frequencies of rhythmic patterns differ significantly between music genres, and thus can be used as a feature in automatic genre classification.

1. INTRODUCTION

Ragtime and jazz are two related genres, both often referred to as “syncopated music”. However, one would not classify Scott Joplin’s “The Entertainer” as jazz, neither would one call Miles Davis a ragtime composer. Yet it is difficult to pinpoint the differences between ragtime and jazz.

From musicological literature, there is evidence that both genres have some characteristic rhythmical patterns. One particular syncopation pattern is considered typical for ragtime (Berlin, 1980), while the clave pattern would be more typical for jazz music (Washburne, 1997). In this paper, we study these patterns by testing musicological hypotheses on data sets of ragtime and jazz music. For this purpose, we introduce JAGAD, a new data set with 252 MIDI files of jazz songs and annotated melody.

We analyze rhythmical patterns taking a corpus-based approach. A corpus-based study is a fast and data-rich way of analyzing rhythmical patterns in many MIDI files. Our research contributes to the fields of Musicology and Music Information Retrieval (MIR). The results can for example be used for automated genre classification, an important task in MIR: if the frequency of occurrences of rhythmical patterns differs significantly between two genres, then this could be used as a new musically meaningful feature, improving music genre classification.

This study builds on earlier research by Volk & de Haas (2013) and Koops et al. (2015). Both studies took a corpus-based approach investigating specific syncopation patterns in ragtime. Volk & de Haas (2013) tested hypotheses about the occurrence of different variations of the same syncopation pattern in ragtime during different periods of the ragtime era (1890 – 1919) and the modern period (1920 – 2012), using a data set of 11591 MIDI files of ragtime pieces. Koops et al. (2015) showed that this syncopation pattern is highly important for the ragtime genre, being among the most frequently used patterns compared to all other patterns.

Contribution. The contribution of the paper is twofold. First, we test and confirm musicological hypotheses about the occurrence of syncopation patterns and the clave pattern in ragtime and jazz melodies using a data-rich approach, thereby contributing to the fields of musicology and Music Information Retrieval (MIR). Second, we introduce JAGAD, a newly collected data set of 252 jazz songs with annotated melody. This data set is not only useful for our current study, but can also be of great use in additional analysis of the jazz genre or in future research on automated melody finding.

2. CHARACTERISTICS OF RAGTIME AND JAZZ

In this section we describe characteristics of ragtime and jazz, focusing on common rhythmical patterns of each genre according to musicology.

2.1 Ragtime

Ragtime was the first black music of the United States that achieved wide commercial popularity (Schaefer & Riedel, 1973). Berlin (1980) researches different theories as to what contributes to ragtime as a genre, such as coon-songs, cakewalk and two-steps. As a possible source of the ragtime rhythm, he mentions dance music of the Caribbean or South America, such as danzas, habaneras and tangos. Hendler (2010) mentions four musical forms that are related to ragtime. The main two roots of ragtime are the quadrille and march. A quadrille is a French contradance - Jelly Roll Morton claimed to have written the famous Tiger Rag from an old quadrille. The march and ragtime are similar in structure. The next related musical element is the cinquillo, which is a syncopated rhythm from the Caribbean. The fourth and last root of ragtime mentioned by Hendler are British popular melodies.

Nowadays, most people associate ragtime with piano music, as they know Scott Joplin’s piano piece “The Entertainer”, which has become widely known after it was used as film score for “The Sting”. Contemporaries of the rag-

ISBN: 978-84-697-2303-6

time period perceived ragtime more as a vocal form. There also exist instrumental ragtime pieces, but piano pieces and vocal songs can be considered the two main instrumentation categories of ragtime (Berlin, 1980).

Berlin (1980) distinguishes three subgroups of piano ragtime: piano renditions of ragtime songs; "ragged" versions of preexisting unsyncopated music and original ragtime compositions. Piano renditions of ragtime songs were not always syncopated. In the second group, existing unsyncopated music, for example marches, popular songs, folk songs and pieces of classical music, was given a syncopated rhythm. But the lion's share of ragtime music comes under the third category: original ragtime compositions for piano. The best-known composers are Scott Joplin, James Scott and Joe Lamb.

In ragtime, the melody usually is the highest pitched line. In case of a piano piece, this is the right hand part of the piece. The accompaniment in the left hand is characterized by stable rhythmical patterns that follow the beat.

2.1.1 The 121 syncopation pattern

Musicologists and ragtime fans have argued that rhythmical patterns and syncopation provide the most distinct features of the ragtime genre. A specific syncopation pattern that is considered important in ragtime by Berlin (1980) is 'short-long-short' or 121 syncopation. The 121 pattern appears as ♪♩♪ in 4/4 meter or as ♬ in 2/4 meter.

Berlin (1980) distinguishes three variants of 121 syncopation in ragtime. The two types that emerge as most important, are untied and tied syncopation. In **untied** syncopation, a pattern does not pass over a bar line and starts on a strong metrical position. In 2/4 meter, the pattern starts either on the first or on the second quarter note position. In 4/4 meter, the pattern starts on the first or third quarter note position. **Tied** syncopation refers to a pattern starting on a weak metrical position. That is, in 2/4 meter it starts at the second or fourth eighth note position and in 4/4 meter it starts at the second or fourth quarter note position. This way, tied syncopation either connects the two halves of a measure, or it connects the second half of a measure to the first half of the next measure. The third pattern is **augmented** syncopation and differs from the other two as it augments the 121 to the length of a complete bar. Figure 1 illustrates these patterns, here in 4/4 meter. On the right of the figure, the patterns are notated in the "onset representation". This is a string with four entries per quarter note. This means that each sixteenth note is represented by one character. If a sixteenth note has an onset, i.e. the start of a note, then the corresponding character is a one. On the other hand, if the sixteenth note has no onset, then the corresponding character is a zero. The dot is a wild-card that matches anything. We use this representation to find patterns in the preprocessed MIDI-files.

2.2 Jazz

The origins of jazz form a contentious subject among musicians, critics and academics. The general belief is that jazz harmonies are based on European practices and that jazz rhythm came from Africa. However, Washburne (1997) and Hendler (2005) throw light on the Caribbean contribution to jazz. The Caribbean influences on jazz are particularly well audible by the rhythm patterns, as we will see in Section 2.2.1.

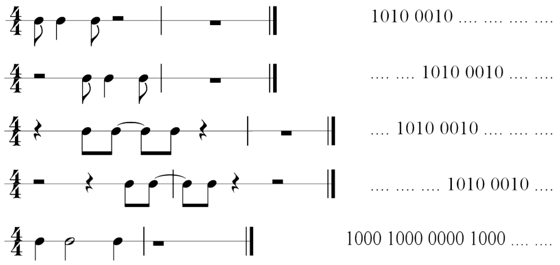


Figure 1: Syncopation patterns. From top to bottom: two variants of untied syncopation, two variants of tied syncopation, augmented syncopation

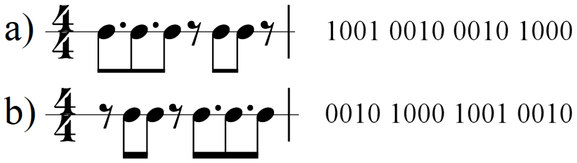


Figure 2: Clave patterns: forward (a) and reverse (b)

Just like ragtime, jazz can be played on a piano. However, it is also commonly played by a jazz ensemble, for example a jazz trio or a bigband.

In contrast to ragtime, it is not always the upper voice that has the melody. For example, the string bass began to be used as a solo instrument, generally from the 1940s on (Kemfeld, 1995).

2.2.1 Clave pattern

In his study of the Caribbean contribution to the genre, Washburne (1997) claims that the clave rhythm is often heard in jazz. This rhythm consists of a syncopated and an unsyncopated part. There are two directions: in forward (or 3-2) clave the syncopated part comes before the unsyncopated part; in reverse (or 2-3) clave it is the other way around. The prototypical patterns of forward and reverse clave are illustrated in Figure 2a and Figure 2b respectively. However, a lot of variations on this pattern are considered as clave. Though Washburne (1997) provides some guidelines, it is not evident to determine if a music phrase or bar is in clave - even for trained listeners.

For this reason,Vurkaç (2012) analyzed the clave and from his research, a new data set was developed, which can be downloaded from the UCI learning repository (Lichman, 2013)[1] . This data set contains 10800 bars in onset notation. Each bar has a label of four bits, indicating the clave direction. 0 0 1 0 and 0 1 0 0 are the clave directions forward and reverse, respectively. The neutral category, labeled 1 0 0 0, refers to patterns that do not detract

[1] https://archive.ics.uci.edu/ml/datasets/Firm-Teacher_Clave-Direction_Classification

Bar onsets	Label	Type
1 1 1 1 0 1 1 1 0 1 1 0 0 1 1 1	**0 0 0 1**	Incoherent
1 1 1 1 0 1 1 1 0 1 1 0 0 1 0 1	**0 0 0 1**	
1 1 0 0 0 1 1 0 0 1 1 0 0 0 1 1	**0 0 0 1**	
1 0 0 1 0 0 1 0 0 0 1 0 1 0 0 1	**0 0 1 0**	Forward
1 0 0 1 0 0 1 0 0 0 1 0 1 0 0 0	**0 0 1 0**	
1 1 1 1 1 0 0 0 1 0 1 0 1 1 1 1	**0 0 1 0**	
0 0 1 0 1 0 0 0 1 0 0 1 0 0 1 0	**0 1 0 0**	Reverse
1 1 1 0 1 0 0 1 1 1 1 1 1 0 1 1	**0 1 0 0**	
0 0 1 0 0 1 1 1 0 1 0 1 0 0 0 1	**0 1 0 0**	
0 1 1 0 0 0 1 0 0 0 0 1 0 0 1 0	**1 0 0 0**	Neutral
0 0 0 1 0 1 1 1 1 1 1 1 0 0 1 1	**1 0 0 0**	
0 0 0 1 1 1 1 1 0 0 0 1 0 0 0 1	**1 0 0 0**	

Table 1: Examples from the UCI clave direction data set

from clave, but do not establish or support any clave direction either. The incoherent category (0 0 0 1) refers to patterns that, in addition to not being in either clave direction, actively oppose the establishment of such. Vurkaç determined these categories based on both double-blind listening tests and informal interviews with four professional master-musicians, as well as decades of studying the music. Table 1 shows three examples per class of the data set. Note that the data set does not give a label for all possible onset combinations: $2^{16} = 65536 > 10800$.

Washburne (1997) points out that the clave pattern is found in several aspects of jazz music: (1) in the rhythmic breaks, for example just before a solo section; (2) in the accompaniment by the rhythm section; (3) in repetitive horn backgrounds or riffs; (4) **in the melody**; and (5) in the phrasing. In this paper, we investigate the occurrence of the clave pattern in the melody of jazz and ragtime pieces.

From his search for samples throughout jazz history, Washburne (1997) observed that some styles incorporate the clave rhythm to a greater extent than others. The pattern is found more often in early jazz (until 1945) than in later styles.

3. A CORPUS BASED STUDY ON RHYTHM PATTERNS IN RAGTIME AND JAZZ

From the characteristics of ragtime and jazz, as described in the previous chapter, three hypotheses about rhythmic patterns in the melodies of ragtime and jazz songs arise.

First, we have seen that tied, untied and augmented 121 syncopation patterns occur frequently in the melody of ragtime songs. We suppose that this is typical for ragtime. That leads to the first hypothesis:

Hypothesis 1 *Tied, untied and augmented 121 syncopation patterns occur more frequently in the melody of ragtime songs than in the melody of jazz songs.*

The second hypothesis is based on the observation by Washburne (1997) that it is easier to find examples in the clave pattern in jazz before 1945 than in jazz of later periods. We wonder if this applies to the frequency of the pattern in the melody too. So the next hypothesis is:

Hypothesis 2 *The clave pattern occurs more frequently in the melody of early jazz pieces (before 1945) than in the melody of later jazz pieces (after 1945).*

In our literature study, we have seen that Caribbean rhythm patterns have influenced both ragtime and jazz. However, we encountered the clave pattern particularly in literature about jazz, and to a lesser degree in books on ragtime. We therefore presume that the clave is more typical for jazz. This leads to our third and final hypothesis:

Hypothesis 3 *The clave pattern occurs more frequently in the melody of jazz pieces than in the melody of ragtime pieces.*

We take a corpus-based approach to test our hypotheses. To this end, we collect two data sets: one with ragtime and one with jazz pieces. Our preprocessing step results in a collection of labeled onsets of 240 jazz and 2579 ragtime songs. The next step is pattern recognition, in which our algorithm calculates for each pattern the proportion of bars in which this pattern occurs. The remainder of this section explains these steps in detail.

3.1 Data set collection

The ragtime data set is a subset of the RAG-collection, as introduced before by Volk & de Haas (2013) and Koops et al. (2015). The complete collection contains 11591 MIDI files of ragtime music. Metadata is added to this data set using a ragtime compendium, consisting of around 15000 ragtime compositions.

Since there were no comparable data sets available for jazz music, we collected a new data set, called JAGAD (Jazz stAndard Gioia Annotated Data set). It should be a collection of songs that is representative for the jazz genre, and of a suitable size: it should consist of sufficient songs to test the significance of aforementioned hypotheses.

The Jazz Standards by Ted Gioia (Gioia, 2012) is a comprehensive guide that lists 252 important jazz compositions. For our data set, we extracted relevant metadata from all songs in this book: the title, composer, lyricist (if applicable) and year of first publication. Subsequently, we located the MIDI files through an extensive web search. As there exist many websites of jazz MIDI files, we were able to find a suitable MIDI for each song. The files have various instrumentations: for example jazz trio, solo piano or big band. In all cases, we chose MIDI's where the melody is clear in at least one channel.

As a next step, the MIDI data has to be prepared for the pattern recognition step. Data preparation consists of melody finding, quantization and filtering of relevant songs.

3.2 Melody finding

Melody finding of the ragtime data set is done automatically, using the skyline algorithm with dip detection, as described by Volk & de Haas (2013). This algorithm takes the highest sounding note when multiple notes sound simultaneously. To overcome that the highest notes from the accompaniment are classified incorrectly as part of the

ISBN: 978-84-697-2303-6

melody at sections where there is no melody, their algorithm sets a lower limit: all notes below the middle C are classified as accompaniment. Also, after performing the skyline algorithm, notes that are characterized by an interval down greater than 9 semitones followed by an interval up greater than 9 semitones are removed. Despite its simplicity, this algorithm works very well for the ragtime genre: evaluating this algorithm on a 435-piece subset of the RAG-collection yields an F-measure of 0.978.

The skyline algorithm is based on the assumption that the melody is (almost) always in the higher-pitched notes. As this is not always the case in jazz (for example in a baritone saxophone solo), the skyline algorithm is unsuitable for extracting the melody in jazz songs. There are no other melody extraction algorithms of which a comparably high accuracy on jazz songs is known. That is why we annotate the melody of the jazz data set by hand. To this end, we use the MuseScore[2] software, which makes it possible to listen to the music and examine a score, generated by the program, at the same time. For each MIDI file, we store the first and last bar number of the melody in a certain channel. This way one or more tuples (barStart, barEnd, channelNr) are associated with each MIDI file. In the following, we only consider the notes that are at that moment part of a melody channel.

The resulting melodies are not always monophonic, so in the "melody channel" it is still possible that two or more notes sound together. In order to extract the rhythm, the melody channel has to be reduced to a monophonic line. For this purpose, we apply the skyline algorithm (introduced as "all mono" by Uitdenbogerd & Zobel (1999)) to the melody notes.

3.3 Quantization

The next step is quantization. A MIDI file consists of note on and note off messages, each starting at a certain time after the previous message. To be able to perform the pattern finding step, we need to translate MIDI timing information into a sixteenth note grid. Our algorithm extracts the note on messages, which correspond to the onset of the note, and quantizes using four bins per quarter note: each onset is assigned to the nearest sixteenth note, as described by Koops et al. (2015). This way, a piece in 4/4 time is represented as a list of 16-character onset strings.

3.4 Filtering relevant songs

Finally, only the relevant songs, that can be matched to the 121 syncopation and clave patterns, are filtered.

In earlier work by Volk & de Haas (2013) and Koops et al. (2015), ragtime songs in 2/2, 2/4 and 2/2 are selected, which have only one meter and start at MIDI tick 0. Furthermore, only rags with a normalized average quantization error up to and including 2% of the corpus are selected. This preprocessing step leads to a list of onsets of 2579 rags.

For the jazz data set, we exclude songs that are not in 4/4 time. This is the case for 12 of the 252 jazz pieces. Omitting these pieces, we have our final preprocessed dataset of onsets of 240 jazz songs.

3.5 Pattern recognition

After creating and preprocessing the data sets, we proceed to the next step: pattern finding, in which we calculate the proportion of bars in which syncopated 121 patterns and the clave patterns appear.

For the tied, untied and augmented 121 syncopation patterns, pattern recognition is straightforward: our algorithm matches each bar (and the first half of the next bar) to each of the patterns and keeps up counters for each of the three 121 syncopation patterns. Then, the results are averaged by dividing each counter by the number of bars of the song.

For matching the clave pattern, a bit more work needs to be done. We use the UCI data set mentioned in Section 2.2.1. For each song, our algorithm counts the number of bars that are in the UCI data set (nrSpecified) and the number of bars that are labeled as forward (nrForward) and reverse (nrReverse) clave. For each song, three real values are calculated, indicating the amount of clave:

$$\text{claveForward} = \frac{\text{nrForward}}{\text{nrSpecified}} \tag{1}$$

$$\text{claveReverse} = \frac{\text{nrReverse}}{\text{nrSpecified}} \tag{2}$$

$$\text{maxClave} = \max(\text{ClaveForward}, \text{ClaveReverse}) \tag{3}$$

4. RESULTS

Having computed the proportion of each pattern per piece, we can now statistically test the three hypotheses introduced in Section 3.

4.1 121 syncopation patterns - ragtime versus jazz

To compare 121 syncopation patterns in the melody of ragtime and jazz music, we analyze ragtime and jazz separately. Table 2 shows the average and median of the proportions of bars in which each variation of the 121 syncopation pattern occurs in ragtime and jazz. In the ragtime data set, on average 15% of the bars contain a tied pattern; in the jazz data set, this is only 1.2%. For the untied patterns, the difference is only a bit smaller: 12% in ragtime as opposed to just 1.5% in jazz. The augmented pattern is not seen very often in the ragtime part (4.0%), but even less in the jazz part: 0.3%.

Note that the syncopation patterns occur so little in jazz that all median values are zero. Figure 3 is a box plot that compares the syncopated patterns in ragtime (blue) and jazz (black).

The jazz and ragtime dataset are very different in size. Therefore, to test the statistical significance of the difference in 121 patterns in the ragtime and jazz dataset, we perform bootstrapping for each of the three 121 syncopation patterns. We compare the 240 jazz pieces with 240 uniformly randomly sampled ragtime pieces, and repeat this

[2] https://musescore.org/

		Tied	Untied	Augmented
Ragtime	Average	0.1497	0.1197	0.0398
Jazz	Average	0.0123	0.0153	0.0030
Ragtime	Median	0.0842	0.0828	0
Jazz	Median	0	0	0

Table 2: Syncopation patterns in ragtime and jazz: average and median of the proportion of bars with the respective pattern, over all songs of the genre.

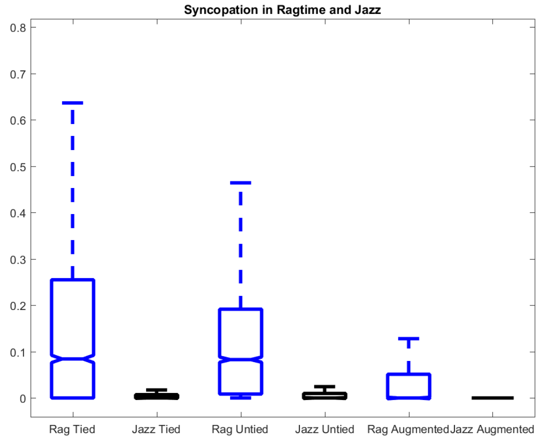


Figure 3: Syncopation in ragtime and jazz

process ten times. Testing for significance using Wilcoxon rank-sum tests for each 121 syncopation pattern, we find a significant difference with $p \ll 0.01$ for every random sample. Therefore, we can conclude that all three 121 syncopation patterns occur significantly more frequently in the melody of ragtime than in jazz pieces, so we accept Hypothesis 1.

4.2 Clave pattern - early versus late jazz

In order to test if the clave pattern occurs more frequently in the melody of early jazz pieces than in the melody of later jazz pieces (Hypothesis 2), we divide the jazz data set into early (year < 1945) and late (year ≥ 1945) jazz. Our data set consists of 148 early jazz pieces and 92 songs of late jazz. We examine the differences between early and late jazz in the proportion of forward clave, reverse clave and the maximum of both directions.

From the results plotted in Figure 4, we see clearly that there is no difference in the use of the clave pattern in the melody of jazz before and after 1945. Forward clave seems to occur a bit more often in early jazz (blue) while reverse clave occurs a bit more frequent in late jazz (black). When we look at the maximum of both directions, the percentage of clave bars is even very similar.

To test for significant differences, we use bootstrapping for each of the three clave pattern variants. We compare the 92 late jazz pieces to a uniform random sample of 92 early jazz pieces. Performing in total 30 Wilcoxon rank-sum tests (for each clave variation and for each random sample) reveals that the differences are not significant: $p > 0.05$ for all samples.

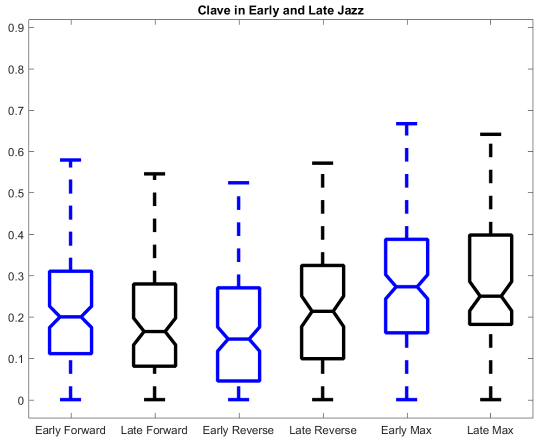


Figure 4: Clave pattern in early (before 1945) and late (from 1945) jazz

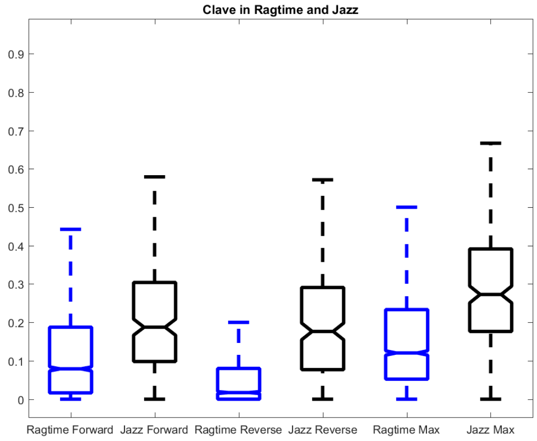


Figure 5: Clave pattern in ragtime and jazz

To conclude, our hypothesis that the clave pattern is more used in early jazz than in late jazz seems not to be acceptable when examining just the melody. We find this outcome somewhat surprising, as it does not correspond to earlier observations by Washburne (1997).

4.3 Clave pattern - ragtime versus jazz

For our final hypothesis, we compare our ragtime to our jazz data set. In Figure 5 we see clearly that all clave pattern directions occur more in jazz than in ragtime. This difference is most obvious when looking at the maximum of both directions: in ragtime, on average 16% of the bars has the largest clave direction; for jazz, this is as much as 29%. Again, we perform bootstrapping and compare the 240 jazz pieces to a uniform random sample of 240 ragtime pieces using Wilcoxon rank-sum tests. We can conclude that these differences are highly significant with $p \ll 0.01$ for all 30 samples. So we accept Hypothesis 3: the clave pattern occurs more frequently in the melody of jazz pieces than in the melody of ragtime pieces.

5. DISCUSSION AND CONCLUSION

In this paper, we investigated the occurrence of several rhythm patterns of ragtime and jazz melodies. We performed a corpus-based study, using computational tools, contributing to the fields of musicology and Music Information Retrieval. As part of our research, we introduced JAGAD, a new data set of 252 jazz MIDI files with annotated melody.

Based on literature research in musicology, we formulated three hypotheses. After performing our corpus-based study, we accepted two hypotheses and rejected a third. The tied, untied and augmented 121 syncopation patterns, as mentioned by Berlin (1980), occur significantly more in ragtime than in jazz. Clave patterns on the other hand occur more frequently in jazz than in ragtime. These outcomes can be used in an automated genre classifier by adding features for the frequencies of syncopation and/or clave patterns: songs with many syncopation patterns are more likely to be ragtime, while songs with many clave patterns are more likely to be jazz. Our last hypothesis, which states that clave patterns are more frequent in early jazz than in late jazz, could not be confirmed.

This research is a next step in the study of typical rhythmic patterns in ragtime and jazz. To investigate if tied and untied syncopation patterns and the clave pattern are truly characteristic for ragtime and jazz respectively, more research on the frequency of these patterns in other genres is needed.

Finally, it would be interesting to examine the frequency of the clave pattern in jazz in other aspects than just the melody. This pattern may occur more often in for example the rhythm or brass section. It is possible that we find a bigger difference between early and late jazz here, which would be in favor of Washburnes hypotheses.

6. REFERENCES

Berlin, E. A. (1980). *Ragtime: A Musical and Cultural History*. Berkeley: University of California Press.

Gioia, T. (2012). *The jazz standards: A guide to the repertoire*. New York: Oxford University Press.

Hendler, M. (2005). *Cubana be Cubana bop*. Graz: Akademische Druck und Verlaganstalt.

Hendler, M. (2010). *Syncopated music: Frühgeschichte des Jazz*. Graz: Akademische Druck und Verlaganstalt.

Kemfeld, B. (1995). *What to Listen for in Jazz*. New Haven: Yale University Press.

Koops, H. V., Volk, A., & de Haas, W. B. (2015). Corpus-based rhythmic pattern analysis of ragtime syncopation. In *Proceedings of the 16th International Society for Music Information Retrieval Conference, ISMIR 2015*, (pp. 483–489). ISMIR press.

Lichman, M. (2013). UCI machine learning repository. [http://archive.ics.uci.edu/ml]. Irvine: University of California, School of Information and Computer Sciences.

Schaefer, W. & Riedel, J. (1973). *The Art of Ragtime*. Baton Rouge: Louisiana State University Press.

Uitdenbogerd, A. & Zobel, J. (1999). Melodic matching techniques for large music databases. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, (pp. 57–66). ACM.

Volk, A. & de Haas, W. B. (2013). A corpus-based study on ragtime syncopation. In *Proceedings of the 14th International Society for Music Information Retrieval Conference, ISMIR 2013*, (pp. 163–168).

Vurkaç, M. (2012). A cross-cultural grammar for temporal harmony in afro-latin musics: Clave, partido-alto and other timelines. *Current Musicology*, (94), 37.

Washburne, C. (1997). The clave of jazz: A caribbean contribution to the rhythmic foundation of an african-american music. *Black Music Research Journal*, *17*(1), 59–80.

ISBN: 978-84-697-2303-6

SCALES IN LITHUANIAN TRADITIONAL FIDDLING: AN ACOUSTICAL STUDY

Rytis Ambrazevičius
Kaunas University of Technology
Lithuanian Academy of Music and Theatre
rytisamb@gmail.com

ABSTRACT

Tunings and intonations in the Lithuanian traditional fiddling are analyzed. 14 archival recordings from 1930s are applied for the acoustical analysis. A huge variety in the individual tunings and intonations is found. Although the fiddles were tuned mostly in the standard way, i.e. in perfect fifths, the fifths are quite "rough". The scales show a strong tendency towards the constituent interval equalizing, or, in other words, certain hybrid scales between the diatonic and equidistant ones are estimated. Concerning the temporal intonations, influence of fiddling motorics and the related influences of bowing force and pitch duration are discussed.

1. INTRODUCTION

Fiddle tunings in the European musical traditions generally follow the standard scheme ("in fifths"), although certain cross-tunings are also frequent (cf. Gurvin, 1968; Anmarkrud, 1992; Ward, 2013). For the present study, more important are the deviations of the tunings and the whole scales from 12TET (the twelve-tone equal temperament). The characteristic deviations leading to the concepts of "blue notes" or "neutral tones" are quite common, for instance, in the Scandinavian and Irish folk music. It is interesting to examine the Lithuanian traditional fiddling in this context.

Fiddle in Lithuanian tradition was mostly used in bands to accompany dances. Older types of the bands contained archaic and newer instruments, such as the lamzdelis (a type of simple recorder), birbynė (a reed pipe), kanklės (a zither-type instrument), dulcimer, fiddle, and basetlė (a bass similar to double bass). The band composition varied among the ethnographic regions. With the introduction of the accordion, in the 19th and early 20th centuries, the bands changed considerably. The newer groups contained almost by necessity an accordion (Petersburg accordion, bandoneon, concertina, etc.) and one or more fiddles. These two instruments still can be documented in fieldwork and are the most popular in contemporary folk dance bands. For instance, more than 700 folk fiddlers are registered in Lithuania (Kirdienė, 2000). A drum and/or basetlė form the rhythm section. Sometimes a clarinet, coronet, and other newcomers would find their way into the band as well.

Similarly as in other traditions, the Lithuanian fiddles were tuned mostly in the standard way, i.e. in perfect fifths[1], although some alternative tunings (scordaturas) are also documented (Kirdienė, 2000, p. 83).

2. SAMPLE AND METHOD

For the present study, I employ the recordings found in the collections Nakienė & Žarskienė, 2003; 2004; and 2005. These collections represent Lithuanian traditional music from different regions as recorded in the 1930s.[2] As a result, we can presume that the instrumental pieces we will analyze reflect relatively old and typical traits of the music.

There are 14 fiddle solo or duo pieces presented in the collections. Four are from Aukštaitija (northeastern Lithuania), three from Dzūkija (southeastern Lithuania), and seven from Suvalkija (southwestern Lithuania). They represent performances by eight fiddlers (Figure 1).[3]

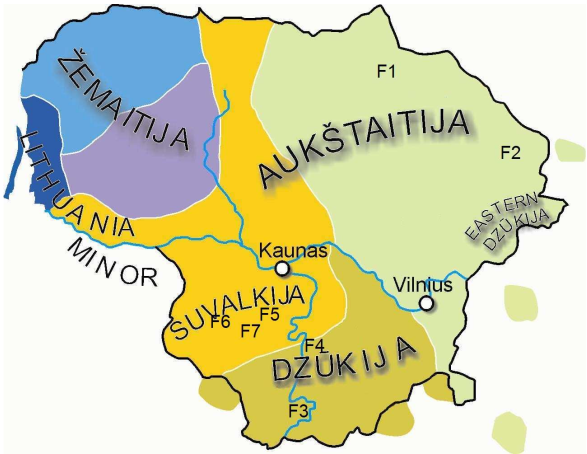


Figure 1. Ethnographic regions of Lithuania. Locations of the fiddlers discussed in the paper (F1, F2…) are indicated.

For the acoustical measurements, software Praat was applied. I determined the tunings of the instruments from their LTAS spectra unless the intonations were noticeably unstable. In the latter case, I performed a more accurate analysis; for instance, I measured pitches note by note or generalized the intonations in particular music contexts.

[1] Later on, we will find that the fifths are quite "rough".

[2] See the Appendix for the fiddler and tune markings, and the detailed information.

[3] There are more fiddle recordings in the collections by Nakienė & Žarskienė, but in groups with other instruments.

3. GENERAL SCALE TRAITS

3.1 Open String Tunings

Regarding the open string tunings, we can only consider the three higher strings (D, A, E) because the lowest one (G) appears clearly in only three performances, T2, T8, and T11. Here I use the standard markings for the four fiddle strings; the actual pitches differ. In every discussed example, the tunings are lower than the standard ones (Figures 2 and 3). Fiddler F4 tuned his fiddle the lowest, with the E string down-tuned by a minor third, or 325 cents. Next was F3 (244–301 cents), F2 (245–253 cents), F5 (220–243 cents), F7 (146–164 cents), F1 (86–96 cents), and, finally, F6 (43–47 cents).

The strings are tuned in rough perfect fifths; any definite general tendency of stretching or shrinking the fifths does not show up. Nevertheless, there are certain tendencies in the tunings by the individual musicians. The fifth D–A is narrowed, while A–E is widened in the four Aukštaitian performances. There is an inverted tendency in F3 tunings. Fiddlers F4, F5, and F7 stretched both fifths. It becomes the case that, no matter how the A is tuned, the ninth D–E is always stretched; it is wider than a 12TET ninth by 9–41 cents, with a median of 24 cents.

3.2 Scales with Reference to a Tonic

Concerning the scales with reference to a tonic, the performances show noticeable scatter. The II degree ranges from 171–205 cents (F7) to 228 cents (F4), with a median of 202 cents. The case of the III degree is more interesting. Except for fiddlers F1 and F4, who intone slightly stretched major thirds, 409–412 cents, the other fiddlers use flattened intonations, down to even 360–380 cents (F6 and F7). The IV degree shows very different intonations, ranging from 475 to 553 cents, with a median of 504 cents. The V and VI degrees in the Suvalkian scales show somewhat less scatter; 681–715 cents for the V degree and 875–928 cents for the VI. However, fiddlers F2, F3, and F4 flatten these degrees noticeably; They intone the V degree at 684–696 cents and 869–886 cents for the VI. Still higher scale degrees appear only in the analyzed Suvalkian performances. The VII scale degree is typically slightly flattened, at around 1086–1098 cents, whereas the VIII degree (octave tonic) is mostly sharpened. Fiddlers F5 and F6 intone at 1209–1238 cents, but F7 flattens to 1187 cents. Importantly, the general tendency of stretching is observed for higher pitches (octave equivalents), as well. A comparison of the different pitches separated by octave gives the range for the mistuned octaves from –7 to +50 cents, with median of +8 cents. The stretching tendency seems to be similar for the pitches below the tonic as well, yet there is too little data to state this confidently. At any rate, low intonations of the 7 degree are clearly pronounced; they range from –98 to –162 cents, with a median of –124 cents.

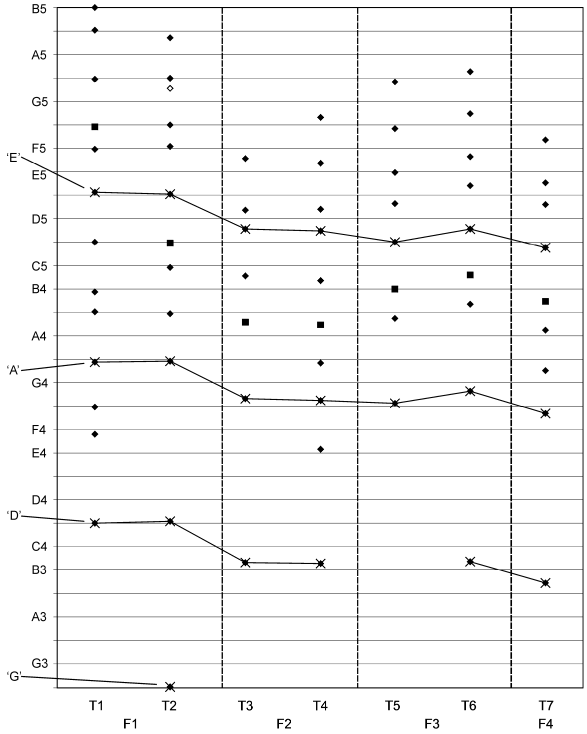


Figure 2. Scales in Aukštaitian (F1, F2) and Dzūkian (F3, F4) fiddle performances. Crossed diamonds denote open strings and squares denote tonics. The white diamonds denote episodic alternative intonation.

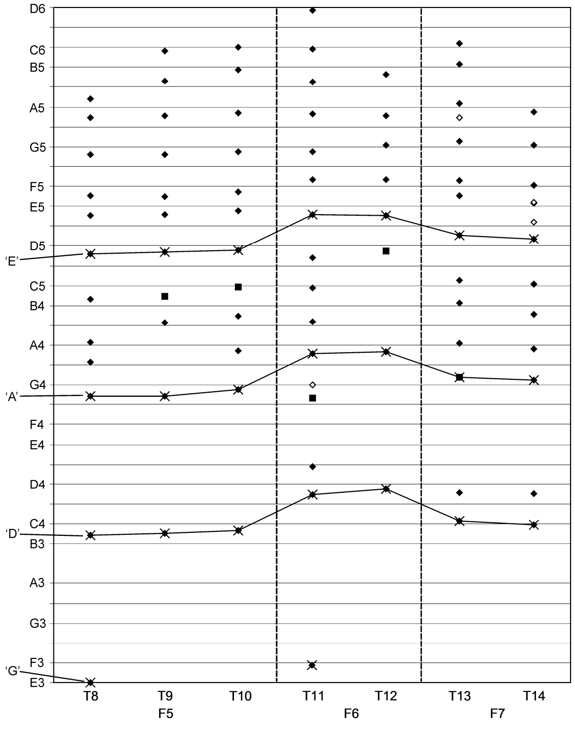


Figure 3. Scales in Suvalkian fiddle performances.

ISBN: 978-84-697-2303-6

3.3 Diatonic Contrast

For the generalization of interval asymmetries in the fiddle scales, we will apply the index of diatonic contrast (DC; Ambrazevičius, 2006, p. 1818). It was introduced as the method of evaluating whether the scale is "more diatonic" or "more in equidistant steps" ("more equitonic"). In other words, the DC defines "by how much the scale is diatonic." The succeeding constituent intervals (i.e. the intervals between the adjacent scale degrees) are pooled into two groups of "small" (d_s) and "large" (d_l) intervals. The following expression for DC is applied:

$$DC = \frac{\bar{d}_l}{\bar{d}_s} - 1 - \frac{2\sum_i \left| d_i - \bar{d}_{(s/l)} \right|}{N\bar{d}_s} \qquad (1)$$

where $N = N_s + N_l$ is the total number of intervals; thus N+1 is the number of scale degrees. The $\bar{d}_{(s/l)}$ means either $\bar{d}_s$ or $\bar{d}_l$, depending on the attribution of d_i.

The formula gives different DC values depending on the grouping. The largest possible value is defined as the actual DC. Defined this way, the diatonic contrast is normalized to 12TET, that is, if the value of the DC is 1, then it means that the corresponding set consists of scale degrees separated by tempered whole tones and semitones. A DC value of 0 means ideal equitonics or equal intervals between degrees (Figure 4).

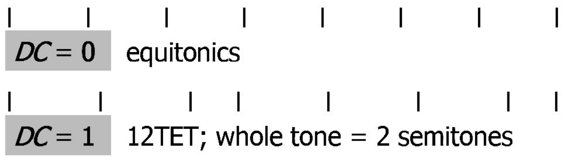


Figure 4. Diatonic contrast.

The more clearly the constituent intervals of a scale cluster into two groups ('small' and 'large' intervals), the larger the DC is. The method of DC is intended to evaluate the overall asymmetry of an intervallic structure. It does not detect the individual differences between the scales, e.g. between the minor and major modes.

F1		F2		F3		F4
T1	T2	T3	T4	T5	T6	T7
.89	.83	1.39	.25	.34	.45	.48

F5			F6		F7	
T8	T9	T10	T11	T12	T13	T14
.80	.30	.47	.05	.13	.55	.13

Table 1. Diatonic contrast for the scales of fiddle performances.

Application of the discussed method to the scales in Figures 2 and 3 results in very diverse values for DC (Table 1).[1] Even the individual fiddlers show quite flexible scaling patterns from the perspective of scale asymmetry. Importantly, majority of the scales are closer to equitonic than to diatonic scale.

4. LOCAL SCALE DYNAMICS

4.1 Flexible Intonation: General Matters

Besides the static aspects of the fiddle scales, there are some interesting phenomena of dynamic intonation. First of all, the performances by the different musicians show different pitch stability. Second, scale degrees sometimes differ considerably in their intonation stability. Compare, for example, the stability of the 5 and II scale degrees realized by open strings with the stability of other scale degrees in the performance T10 (Figure 6). Consider the flexible intonation in this piece, especially of the VII degree (Figure 7). There are several causes of and aspects to the differences. Obviously, the left hand's position and fingering matter. Therefore, naturally, the pitches of the open strings are expected to be intoned the most steadily.

4.2 Influence of Bowing Force

Even the pitches of the open strings, as well as the rest, are slightly affected by bowing force. The force phenomenon is observed, for instance, in T11; it results in a double peak for the tonic in the LTAS (Figure 9). This degree (note G4 in Figure 8) is more accented in the first occurrence of the sequence G4–B4–G4–B4 (see the first G4, with an arrow, in Figure 8) and played more smoothly in the second occurrence. Therefore, the first pitch is higher than the second one by 65 cents, on average. It is reasonable to consider the lower pitch as performed in a "normal" way and, hence, as the "correct" pitch of the scale.

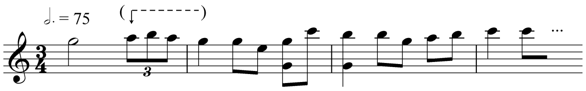


Figure 5. Transcription of the beginning of T10. For detailed intonations, see Figure 7.

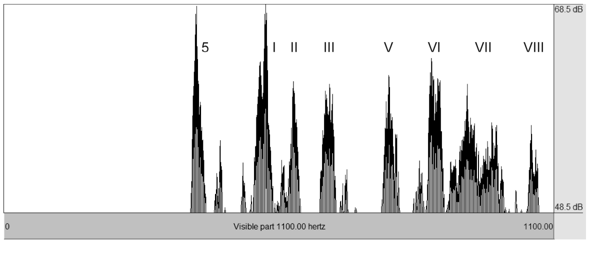


Figure 6. LTAS of T10. The most prominent scale degrees are marked.

[1] Of course, only the intervals between adjacent scale pitches are considered. The gaps in the scales (i.e. if some of the pitches are not used in the tune) are not considered.

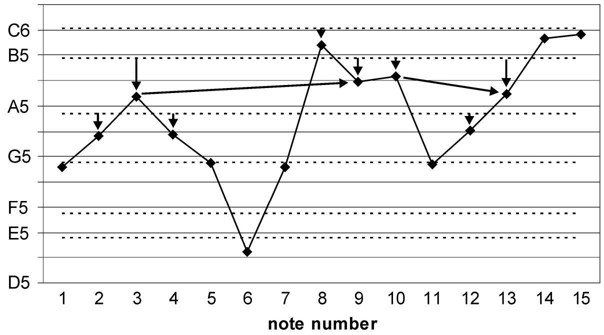


Figure 7. Pitch track of the first phrase of T10; see the transcription in Figure 5. Dotted lines indicate the measured average pitches of the scale degrees. Vertical arrows indicate the local intonation of the scale degrees. Long tilting arrows indicate the change of the VII degree intonation.

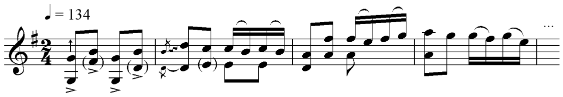


Figure 8. Transcription of the beginning of T11. Detailed intonations not depicted.

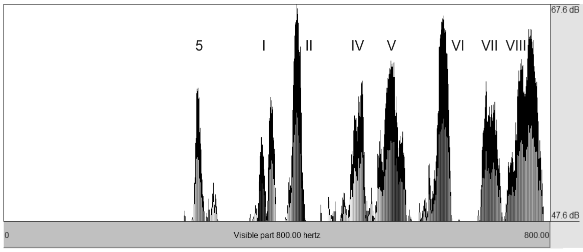


Figure 9. LTAS of T11. The most prominent scale degrees are marked.

4.3 Influence of Pitch Duration

Because of the different motoric patterns, pitches produced with one fingering may differ from the pitches produced with another fingering. An additional factor is pitch duration. See T3, for instance. The V degree (the highest note, F5, in Figure 10) is performed on the A string with an uncomfortably high position; the notes are short, and the finger has to move quickly to its position. Therefore, shorter notes tend to be performed with slightly flattened pitches (Figure 11). When interpreting the musical scale, it is reasonable to accept the longer pitches, or, in other words, the pitches not affected by the technical motoric constraints, as the "correct" ones.[2]

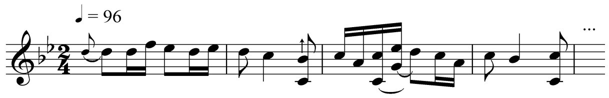


Figure 10. Transcription of the beginning of T3. Detailed intonations not depicted.

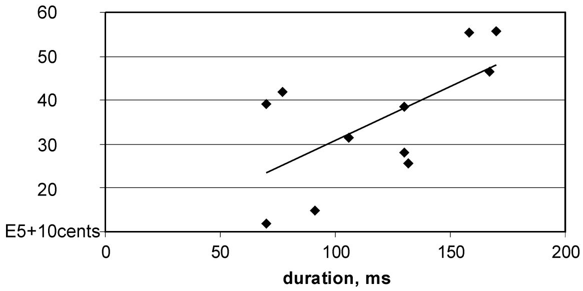


Figure 11. Fiddle piece T3. Dependence of the intonations of the V degree on pitch duration.

It is not clear whether the changeable intonation of the tonic (Figure 12) could be explained by similar technical causes or, rather, by certain intentional "chromaticisms." The latter option seems to be less possible, as the two pitch zones are not distinctly separated. We suggest that the "correct" pitch of the tonic is the lower one, as it is realized by longer notes and in more relevant metric positions.

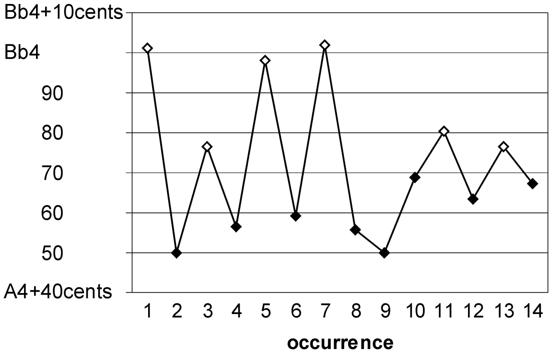


Figure 12. Fiddle piece T3. Intonations of the first 14 occurrences of the tonic. White diamonds show B♭ 4 in the second measure (Figure 10) and in similar sequences. Black diamonds show B♭ 4 in the fourth measure and in similar sequences.

There are more instances of local intonations that depend on the musical contexts in the discussed performances. In T13, the IX degree (the octave counterpart of the II degree) is intoned sharper in the first part of the tune and flatter in the second part, after modulation. The difference is even about 70 cents. Again, it is reasonable to consider the longer pitches as the "correct" ones. In this case,

[2] We possibly encounter cases when there are no occurrences of the analyzed scale degree that are not affected by the constraints. Then we consider the occurrences that are affected the least bit possible, yet we should keep in mind these are still not "correct."

those are the higher pitches. They are denoted by the corresponding black diamonds in Figure 3, and the white diamonds denote the lower pitches. In T14, the VII degree shows even three characteristic intonations in different contexts (Figure 3). Performance T14 is also interesting for a “sliding” technique: Fiddler F7 typically applies slight positional shifts while producing individual pitches (Figure 13).

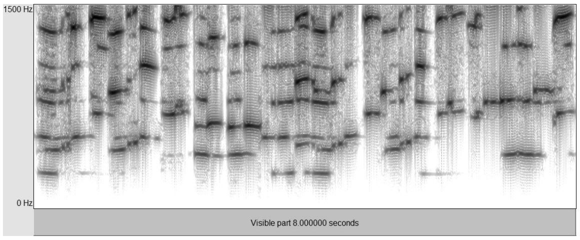


Figure 13. T14. Fragment of spectrogram; the first period of the piece.

5. DISCUSSION

The analysis showed a huge variety in the individual tunings and intonations. However, some generalizations can still be made. First, the fiddles were tuned more or less (up to three semitones) lower than prescribed by the standard G-D-A-E scheme. One reason could be the fact that the fiddlers used to accommodate to accordions while playing in the band (Kirdienė, 2000, p. 83). The lowered fiddle tuning has been observed in other traditions as well (cf. Ward, 2013). Second, and more interesting, the tendency toward scale stretching is noticeable. Third, while diatonic major patterns seemingly prevail, the significant flattening of the VII degree and its octave equivalent 7 degree, as well as the flattening of the III degree in some cases, diminish the diatonic contrast. These quasi-equitonic tendencies are most clearly pronounced in the performances of fiddlers F3 and, especially, F6. For instance, the scale fragment I–VIII in T11 shows the following interval sequence: 221–159–173–150–219–175–142 cents. The conclusion about the tendency towards the constituent interval equalizing is supported by the evaluations of diatonic contrast.

Actually the hybrid scales between the diatonic and equidistant ones make no surprise; various manifestations of such scales were abundantly traced in the Lithuanian traditional instrumental and vocal music (Ambrazevičius, Budrys, & Višnevska 2015), and elsewhere (cf. Grainger, 1908–1909; Sevåg, 1974). It is very likely that such scales reflect relics of certain archaic musical thinking. Incidentally, the deviations in these scales should not be confused with quarter-tones characteristic, for example, of Middle Eastern maqam or Occidental quarter-tone (or microtonal) music, such as Ligeti. The deviations examined here neither are steady, nor intended as quarter-tones.

The discussed scales are not static but rather dynamic, that is, the scale pitches are not equally intoned in the course of a performance. In addition to possible peculiar phenomena of musical thinking, here motorics of fiddling could be at work. The motorics determines more and less comfortable fingerings; fast passages produce less “reliable” pitches compared to the slow movements. All this results in a performance “noise”, but supplemented with noticeable systematic tendencies of intonation.

6. APPENDIX

The markings for fiddlers and tunes used in the body text are explained. The tunes T1–T4, T5–T7, and T8–T14 come, respectively, from Aukštaitija, Dzūkija, and Suvalkija regions, and are published in Nakienė & Žarskienė, 2004; 2005; and 2003. The numbers in angle brackets indicate the running numbers of the tunes in the publications. LTRF is abbreviation for *Lietuvių tautosakos rankraštyno fonoteka* (Audiorecords at the Archives of Lithuanian folklore, Institute of Lithuanian Literature and Folklore).

F1: Mikas Marciukas (Papilys, Biržai county) and Kazys Latvėnas (Vieščiūnai, Vabalninkas township, Biržai county). Records of 1935.

F2: Juozas Gudėnas (55, Bajorai, Salakas township, Zarasai county; b. in Bajorai, Daugėliškis township, Švenčionys county). Records of 1939.

F3: Kostantas Lukoševičius (24, Guobiniai, Leipalingis township, Seinai county). Records of 1935.

F4: Julius Kapka (72, Makniūnai, Nemunaitis township, Alytus county). Records of 1935.

F5: Juozas Radzevičius (70, Žarsta, Klebiškis township, Marijampolė county). Records of 1936.

F6: Jurgis Byla (71, Vizgirdai, Paežeriai township, Vilkaviškis county). Records of 1937.

F7: Jurgis Gudynas (75, Veselava, Javaravas township, Marijampolė county). Records of 1937.

T1: Pandėlio polka [The Pandėlys Polka]. LTRF disc 60(2) [57].

T2: Kuzma (Tuina) polka. LTRF disc 60(4) [75].

T3: Ožiukas [The Kid Goat]. LTRF disc 1169(3) [70].

T4: Rātasai [The Round]. LTRF disc 1170(2) [71].

T5: Cikano polka [The Cikanas Polka]. LTRF disc 115(9) [40].

T6: Zantos polka [The Zanta Polka]. LTRF disc 115(10) [6].

T7: Šilinė polka [The Pine-Wood Polka]. LTRF disc 305(1) [36].

T8: Polka. LTRF disc 443(1) [12].

T9: Polka. LTRF disc 443(2) [14].

T10: Mazurpolkė „Aukštoji“ [Mazurpolka ‘The Highest’]. LTRF disc 443(6) [15].

T11: Polka. LTRF disc 651(1) [22].

T12: Vestuvinis maršas išleidžiant [Wedding March at Parting]. LTRF disc 652(2) [26].

T13: Polka. LTRF disc 645(1) [23].

T14: Polka. LTRF disc 644(4) [38].

7. REFERENCES

Ambrazevičius, R. (2006). Pseudo-Greek modes in traditional music as result of misperception. In *ICMPC9. Proceedings of the 9th International Conference on Music Perception and Cognition. 6th Triennial Conference of the European Society for the Cognitive Sciences of Music. Alma Mater Studiorum University of Bologna, Italy, August 22–26, 2006* (CD). Bologna: Bononia University Press (pp. 1817–1822).

Ambrazevičius, R., Budrys, R., & Višnevska, I. (2015). *Scales in Lithuanian traditional music: Acoustics, cognition, and contexts*. Kaunas: Kauno technologijos universitetas, Green Prints.

Anmarkrud, B. (1992). *De ulike felestillene i hardingfeletradisjonene [Various fiddling styles in the Hardanger fiddle tradition]*. Rauland, Norsk Folkemusikklag.

Grainger, P. (1908–1909). Collecting with the phonograph. *Journal of the Folk Song Society, 3*, 147-242.

Gurvin, O. (1968). The Harding fiddle. *Studia Musicologica Norvegica, 1*, 9-20.

Kirdienė, G. (2000). *Smuikas ir smuikavimas lietuvių etninėje kultūroje [Fiddle and fiddling in Lithuanian ethnoculture]*. Vilnius: Kronta.

Nakienė, A. & Žarskienė, R., eds. (2003). *Suvalkijos dainos ir muzika. 1935–1939 metų fonografo įrašai / Songs and music from Suvalkija. Phonograph records of 1935–1939*. Vilnius: Lietuvių literatūros ir tautosakos institutas.

Nakienė, A. & Žarskienė, R., eds. (2004). *Aukštaitijos dainos, sutartinės ir instrumentinė muzika. 1935–1941 metų fonografo įrašai / Songs, Sutartinės and music from Aukštaitija. Phonograph records of 1935–1941.* Vilnius: Lietuvių literatūros ir tautosakos institutas.

Nakienė, A. & Žarskienė, R., eds. (2005). *Dzūkijos dainos ir muzika. 1935–1941 metų fonografo įrašai / Songs and music from Dzūkija. Phonograph records of 1935–1941*. Vilnius: Lietuvių literatūros ir tautosakos.

Sevåg, R. (1974). Neutral tones and the problem of mode in Norwegian folk music. In G. Hilleström, ed. *Studia Instrumentorum Musicae Popularis III* (pp. 207–213). Stockholm: Musikhistoriska Museet.

Ward, C. (2013). Scordatura in the Irish traditional fiddle music of Longford and South Leitrim. *The Musicology Review, 8*, 109-129.

MICROINTERVAL MODALITY IN TRADITIONAL IRISH MUSIC – AN EMPIRICAL APPROACH

Dr Ryan Molloy
National University of Ireland,
Maynooth
ryan.molloy@nuim.ie

1. INTRODUCTION

The collection and publication of tunes in Irish traditional music since the 17th century has been well documented; however, one facet of Irish traditional music has been largely unexplored by ethnomusicologists as an apparent result of the collection process, namely the plausibility of microinterval modality in Irish traditional music. Very little can be inferred about this from the written history of Irish music up to the 20th century, but certain glimpses into this 'lost music' are provided by some of the first recordings of Irish music in the early part of the 20th century, as well as a text by Rev. Richard Henebry. In these sources, it is apparent that the use of non-tempered scales is not as uncommon as is currently thought. This oral paper presents preliminary evidence to support some of these earlier findings through the examination of a small selection of archival recordings using basic empirical methods to analyse microintervallic content.

2. METHODOLOGY

Reccordings of singer Brigid Tunney, fiddler Bobby Casey and piper Patsy Touhey are selected initially through aural determination of microintervallic content. The pitch content of the selcted recordings is then analysed using the software *Praat* to create melographs which can then be used to both visually determine microintervals and the precise pitch involved. In addition, the data obtained can be used to create 'pitch-centre plots' that demonstrate the variation of specific pitches in a particular melody. From these, preliminary evidence for the systematic usage of microintervals in traditioanl performance can be gathered. The preliminary results suggest a reappraisal of the melodic content of Irish traditional music and the necessity of a wider survey, as well as providing reasons for the long-standing tradition of monophony in early Irish music.

3. RESULTS

This cursory study reveals that in each of the tracks examined there is a prevalent variance in the intonation of fourths and sevenths in the scale, which were frequently found a quartertone away from their equally tempered equivalents. Other scale degrees showed smaller variation by comparison, although a quarter-flattened third is prevalent in Casey's performance (remarkably similar to a 'neutral' scale). These observations tie in with Henebry's frequent reference to the 'advanced' fourths and sevenths and lends some weight to his classification of 'Irish scales'. However, there is insufficient evidence here to proffer any agreement with Henebry's assertion that the position of these notes varies depending on the direction of the melody (ascent vs. descent).

4. DISCUSSION

The question remains as to why this variation is present at all. It would be very easy to proffer an answer along the lines of Henebry in saying that their very existence is reason enough in itself. I tentatively suggest, in agreement with the eminent Irish music scholar Breandán Breathnach, that the answer lies in the question of 'gapped scales'. Two of the pieces examined here show a clear usage of a pentatonic scale, but not in the usual sense: (thinking in terms of G as prime) *The Wee Weaver* and *The Munster Gimlet* are based on the notes G-A-C-D-F, with C and F being variable between quarter-sharp and fully sharp. The usual pentatonic scale, G-A-B-D-E, lacks these variable notes and it is interesting to note that the quartertones fall precisely between the two 'gaps' in this scale. This may support the primordialism of the latter scale in Irish music, with the more recent addition of variable fourth and sevenths exploratively creating new means of expression. Alternatively, there may be some other grounding for the preference of the G-A-C-D-F pentatonic scale over G-A-B-D-E in older Irish music. Further examination of this is beyond the scope of this discussion. One explanation for the variance of thirds and sevenths could lie in the possibility of 'hybrid modes'.

5. CONCLUSION

The results discussed in this paper are significant in both the preservation of Irish musical heritage and cultivation of a new direction in Irish contemporary music, tackling head-on the divide between contemporary Irish art music and traditional music. This paper discusses briefly the schism between the two main branches in Irish contemporary music and how the possibility of microinterval modality in Irish traditional music can act as a bridge between them.

Oral session 2

Folk Music Analysis
MÁLAGA, Spain
14-16 June 2017

AN IMMUNE-INSPIRED COMPOSITIONAL TOOL FOR COMPUTER-AIDED MUSICAL ORCHESTRATION

Marcelo Caetano mcaetano@ic.uma.es

Isabel Barbancho ibp@ic.uma.es

Lorenzo Tardón lorenzo@ic.uma.es

Aplicación de las Tecnologías de la Información y Comunicaciones (ATIC)

Universidad de Málaga (UMA)

ABSTRACT

The aim of computer-aided musical orchestration (CAMO) is to find a combination of musical instrument sounds that approximates a target sound or a desired timbral quality. The difficulty arises from the complexity of timbre perception and the combinatorial explosion of all possible instrument mixtures. The state of the art uses Genetic algorithms (GAs) to explore the vast space of possible instrument combinations with a fitness function that encodes timbral similarity between the candidate instrument combinations and the target sound. However, GAs tend to lose diversity during the search, resulting in only one orchestration. In this work, we propose to use an artificial immune system (AIS) called opt-aiNet to search for candidate orchestrations because opt-aiNet returns multiple orchestrations that are all similar to the target yet different from one another. Diversity results in the ability to provide the composer with multiple choices when orchestrating a sound instead of searching for one solution constrained by choices defined a priori. Therefore, diversity can expand the creative possibilities of CAMO beyond what the composer initially imagined.

1. INTRODUCTION

Orchestration refers to composing music for an orchestra (Kendall 1993). Initially, orchestration was simply the assignment of instruments to pre-composed parts of the score, which was dictated largely by availability of resources such as the number and type of instruments available (Handelman 2012, Kendall 1993). Later on, composers started regarding orchestration as an integral part of the compositional process whereby the musical ideas themselves are expressed (Rose 2009). Compositional experimentation in orchestration arises from the increasing tendency to specify instrument combinations to achieve desired effects, resulting in the contemporary use of timbral combinations (McAdams 1995, Rose 2009, Abreu 2016). The development of computational tools that aid the composer in exploring the virtually infinite possibilities resulting from the combinations of musical instruments gave rise to computer-aided musical orchestration (CAMO) (Carpentier 2006,2007,2010a,b, Hummel 2005, Psenicka 2003, Rose 2009). Most of these tools rely on searching for combinations of musical instrument sounds from pre-recorded datasets to approximate a given target sound. Early works (Hummel 2005, Psenicka 2003, Rose 2009) resorted to spectral analysis followed by subtractive spectral matching.

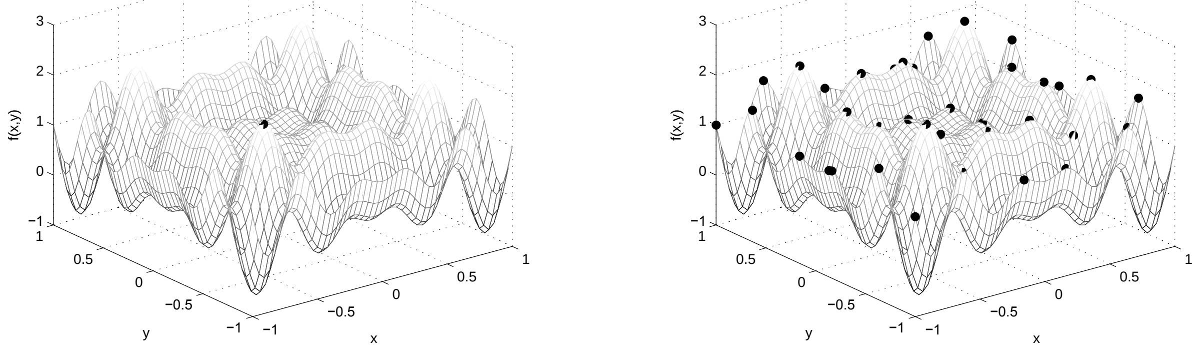
To overcome the drawbacks of spectral matching, Carpentier and collaborators (Carpentier 2006,2007,2010a,b, Tardieu 2007) search for a combination of musical instrument sounds whose timbral features best match those of the target sound. This approach requires a model of timbre perception to describe the timbre of isolated sounds, a method to estimate the timbral result of an instrument combination, and a measure of timbre similarity to compare the combinations and the target. Multidimensional scaling (MDS) of perceptual dissimilarity ratings (McAdams 1995, Caclin 2005) provides a set of auditory correlates of timbre perception that are widely used to model timbre perception of isolated musical instrument sounds. MDS spaces are obtained by equating distance measures to timbral (dis)similarity ratings.

Preliminary work (Abreu 2016) used an artificial immune system (AIS) called opt-aiNet (de Castro 2002) to search for multiple combinations of musical instrument sounds whose timbral features match those of the target sound. Inspired by immunological principles, opt-aiNet returns multiple good quality solutions in parallel while preserving diversity. The intrinsic property of maintenance of diversity allows opt-aiNet to return all the optima (global and local) of the fitness function being optimized upon convergence, as shown in Fig.1 b). Fig.1 illustrates a two-dimensional fitness function to be optimized by finding the peaks (i.e., points where the function has maximum amplitude). Fig.1 a) illustrates the mono-modal optimization property of GAs, which typically converge to a unique solution represented as a black dot on a peak. In turn, Fig. 1 b) illustrates the multimodal ability of the AIS opt-aiNet, capable of returning all optima of the function within the region of interest.

The application of opt-aiNet in CAMO gave rise to Immune Orchestra (Abreu 2016). The preliminary results suggest that the property of maintenance of diversity translates as orchestrations that are all similar to the target yet different from one another. Thus Immune Orchestra can provide the composer with multiple choices when orchestrating a sound instead of searching for one solution constrained by choices defined a priori (Carpentier 2010a). Consequently, Immune Orchestra has the potential to expand the creative possibilities of CAMO beyond what the composer initially imagined.

We are currently improving Immune Orchestra to ensure maximum diversity of the proposed orchestrations both objectively and subjectively. We will present the results, compare with the state of the art, and discuss potential improvements for future work.

ISBN: 978-84-697-2303-6



a) Standard Genetic Algorithm (GA)

b) Artificial Immune System (AIS)

Fig.1: Comparison between mono-modal (GA) and multi-modal (AIS) optimization. The AIS used is opt-aiNet.

2. REFERENCES

J. Abreu, M. Caetano, R. Penha (2016) Computer-Aided Musical Orchestration Using an Artificial Immune System. In: C. Johnson, V. Ciecielski, J. Correia, P. Machado (Eds.) Evolutionary and Biologically Inspired Music, Sound, Art and Design. Springer International Publishing, vol. 9596, pp. 1-16.

A. Caclin, S. McAdams, B. Smith, S. Winsberg (2005) Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. *Journal of the Acoustical Society of America* 118(1), pp. 471–482.

G. Carpentier, D. Tardieu, G. Assayag, X. Rodet (2006) E. Saint-James. Imitative and generative orchestrations using pre-analysed sound databases. In: *Proceedings of the Sound and Music Computing Conference*, pp. 115–122.

G. Carpentier, D.Tardieu, G. Assayag, X. Rodet, E. Saint-James (2007) An evolutionary approach to computer-aided orchestration. In: Giacobini, M. (ed.) EvoWorkshops 2007. LNCS, vol. 4448, pp. 488–497.

G. Carpentier, G. Assayag, E. Saint-James (2010a) Solving the musical orchestration problem using multiobjective constrained optimization with a genetic local search approach. *Journal of Heuristics* 16(5), pp. 681–714.

G. Carpentier, D. Tardieu, J. Harvey, G. Assayag, E. Saint-James (2010b) Predicting timbre features of instrument sound combinations: application to automatic orchestration. *Journal of New Music Research* 39(1), pp. 47–61.

L. N. de Castro, J. Timmis (2002) An artificial immune network for multimodal function optimization. In: *Proceedings of the 2002 Congress on Evolutionary Computation*, vol. 1, pp. 699–704.

E. Handelman, A. Sigler, D. Donna (2012) Automatic orchestration for automatic composition. In: *1st International Workshop on Musical Metacreation* (MUME 2012), pp. 43–48.

T. Hummel. Simulation of human voice timbre by orchestration of acoustic music instruments (2005) In: *Proceedings of the International Computer Music Conference* (ICMC), pp. 185.

R. A. Kendall, E. C. Carterette (1993) Identification and blend of timbres as a basis for orchestration. *Contemp. Music Rev.* 9(1–2), pp. 51–67.

S. McAdams, S. Winsberg, S. Donnadieu, G. DeSoete, J. Krimphoff (1995) Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychol. Res.* 58(3), pp. 177–192.

D. Psenicka. SPORCH: an algorithm for orchestration based on spectral analyses of recorded sounds (2003) In: Proceedings of International Computer Music Conference (ICMC), pp. 184.

F. Rose, J. E. Hetrik (2009) Enhancing orchestration technique via spectrally based linear algebra methods. *Computer Music Journal* 33(1), pp. 32–41.

D. Tardieu, X. Rodet (2007) An instrument timbre model for computer aided orchestration. In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 347–350.

3. ACKNOWLEDGEMENTS

This work has been funded by Ministerio de Economía y Competitividad of the Spanish Government under Project No. TIN2016-75866-C3-2-R. This work has been done at Universidad de Málaga, Campus de Excelencia Internacional Andalucía Tech.

A SEMI-AUTOMATIC METHOD TO PRODUCE SINGABLE MELODIES FOR THE LOST CHANT OF THE MOZARABIC RITE

Geert Maessen
independent scholar
gmaessen@xs4all.nl

Peter van Kranenburg
Meertens Instituut
peter.van.kranenburg@meertens.knaw.nl

ABSTRACT

The Mozarabic rite provided the dominant context for Christian worship on the Iberian Peninsula and Southern France from the sixth till eleventh centuries. Over 5,000 chants of the Mozarabic rite are preserved in neumatic contour notation. Since pitch-readable notation only became in use in the eleventh century and hardly any Mozarabic chant was found in such notation, scholars believe that most Mozarabic melodies are irretrievably lost. Based on similarities between the chant of Mozarabic and other rites, this paper presents a method for the computational composition of melodies agreeing in all detail with our knowledge of the early Mozarabic neumatic notation. We first describe how we came to look for such a method. Then we give a detailed description of the eight steps of the method. Finally, we propose objective criteria that supposedly are indicative for the authenticity of our compositions, we restate our goals, and refer to several sound examples on the internet.

1. INTRODUCTION

The study of medieval chant-repertoires is of great importance for our understanding of the transition from primarily oral musical cultures to the written and notated history of Western music. At least five medieval chant-repertoires (partially) survive in pitch readable notation from the eleventh and twelfth centuries: Gregorian, Milanese, Old-Roman, Beneventan and Mozarabic chant. Their histories are closely related and go back to times long before the eleventh and twelfth centuries (Hiley, 1993; Fernández de la Cuesta, 2013). Two of these repertoires are also preserved in tenth-century neumatic notation: Gregorian and Mozarabic chant (see Table 1). The Mozarabic rite and its chant were officially abolished in 1085 and replaced by the Roman rite with its Gregorian chant. In Toledo, however, Mozarabic chant survived orally until it was partly notated in sixteenth century mensural notation. In these pitch-specific notations hardly any correspondence can be found with the early neumatic notation.

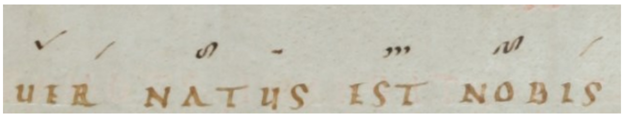


Figure 1. Beginning of the introit *Puer natus,* CH-SGs 339, St. Gall, Switzerland, 980-1000 (initial *P* omitted)

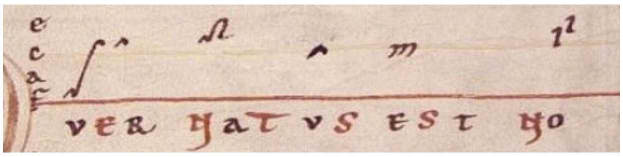


Figure 2. Beginning of the introit *Puer natus,* A-Gu 807, St. Florian, Austria, XII c. (initial *P* omitted). The two horizontal lines respectively represent the f (lower line) and the c' (upper line).

Neumatic notation was meant as a memory aid. It consists of a sequence of symbols written above the text, indicating the contour of the melody. For example, the first neume in Figure 1 refers to an ascending interval of two notes. The second neume represents a single note. The

	neumatic notation		correspondence	pitch notation		database
	century	*chants*		*century*	*chants*	*C*
GRE: Gregorian chant	X-XI	4,000	> 99 %	XI-XII	10,000	281
MIL: Milanese chant	-		-	XI-XII	3,000	167
ROM: Old Roman chant	-		-	XI-XII	3,000	141
BEN: Beneventan chant	-		-	XI-XII	200	51
MOZ: Mozarabic chant	X-XI	5,000	< 1 %	XI-XVI	500	149

Table 1. Estimation of the number of chants in five traditions, the correspondence between the two types of notation, and the number of chants in our database *C*.

third neume represents an ascending interval followed by a descending interval. In the early sources, the exact sizes of the intervals are not indicated. The historic performer knew the melody by heart.[1] For us, the only way to obtain knowledge of the historical melodies is by consulting sources from later date that contain pitch-specific notation for corresponding chants. We can find these corresponding chants by comparing liturgical assignment (feast and function) and texts of the chants. For example, in Gregorian chant, the mass of Christmas Day starts with the introit *Puer natus.* Comparing unpitched tenth-century neumes of this chant (see Figure 1) with corresponding twelfth-century pitches (see Figure 2), we can see a perfect correspondence in musical detail. Both introits (most likely) refer to the same melody (`gd'-d' d'e'd'-c' c'c'c' d'c'e'd'`, on *Pu-er na-tus est no-*).

However, virtually all Mozarabic chants preserved in early, unpitched neumatic notation do not correspond to their pitch-readable counterparts, if these exist at all. Figure 3 shows the Mozarabic *Parvulus natus,* the sacrificium (offertory) for the mass of Christmas Day. Just to mention one striking difference, we can observe different numbers of notes in the two versions. The neume above the first syllable of *Parbulus* in León indicates a single note. Above the second syllable we see three notes: a single note is followed by a *clivis,* an ascending note followed by a descending note. Conform Rojo and Prado (1929) and González-Barrionuevo (2015) León thus shows 1-3-6, 1-2, 4 and 13-2 notes on the first four words, while Toledo shows 1-1-1, 3-3, 2 and 4-9.

Because of this lack of correspondence, the vast majority of Mozarabic melodies are unknown. Therefore, until recently, Mozarabic chant did not receive much scholarly attention. The contrast with Gregorian chant is reflected in the correspondence figures in Table 1. The availability of melodies would greatly improve our access to the lost tradition. As the exact reconstruction of the musical past obviously seems not achievable, in this paper we aim to construct historically informed, *singable* melodies that correspond with the neumatic contour notation of the early manuscripts. This, at least, will render the repertoire performable for contemporary ensembles.

Our first attempt was to compose chants manually ourselves. We put the results of this to the test by providing them to the ensemble Gregoriana Amsterdam during a regular rehearsal, without telling the source of the chants. The singers, however, did not accept the chants. They even guessed themselves that these were new compositions before they were informed about it. The melodies apparently disagreed too much from the styles familiar to them.

Next, we asked several professional composers to make melodies that agree with the early contour notations, in order to rehearse and perform these. One of them made several different compositions in different modes for several chants. We rehearsed and performed a selection of these (Swaan, 2012). A second composer explored her artistry in microtonal directions (Driessen, 2013). In all cases, however, the chants the composers produced, stylistically seemed not to correspond to our knowledge of the five traditional styles of Table 1. We became convinced that inviting modern composers was not the best option to recreate something of the lost Mozarabic music.

We set out to design a more objective method to find pitches that match the neumatic contour notation. We know that the five pitch-readable chant repertoires are interrelated (Hiley, 1993; Levy, 1998). Therefore, it is plausible to employ the melodic material from these traditions for our purpose. Our current approach is to automatically search a database of digitized chants from pitch-readable sources in order to retrieve a melody that matches the neumatic contour notation of a lost melody as much as possible. This procedure renders the lost melodies singable again using stylistically related historic melodic material.

This paper presents our method. We first discuss the representations we use. Then, the central steps of the method are presented. We propose a scoring mechanism to evalutate the authenticity of the retrieved melody and conclude with some examples performed by Gregoriana Amsterdam on YouTube.

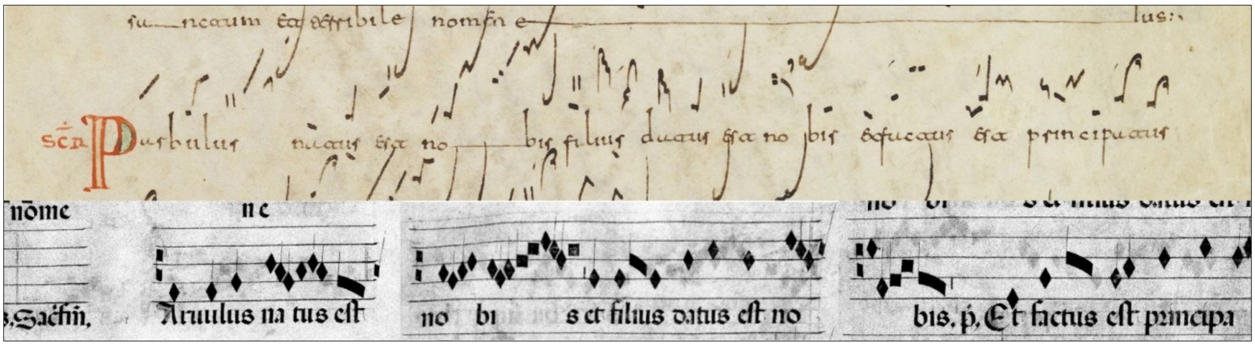

scr. Parbulus natus est nobis filius datus est nobis et factus est principatus
nōme
ne
Sacrifi.
Aruulus na tus est
no bi s et filius datus est no
bis. p. Et factus est principa

Figure 3. Beginning of *Parvulus natus,* top, E-L 8, León, early X c; bottom, E-Tc Cantoral I, Toledo, early XVI c.

[1] In fact, the historic performer did not perform from notation at all. In the best case notation was only used as a reference (see Hiley, 1993).

2. METHOD

The method we propose consists of the following eight steps:

1. Represent the neumes as sequence of contour letters *t*;
2. Construct a database *C* with pitched melodies;
3. Divide *t* into phrases;
4. Find a matching source melody *s* for *t* in *C*;
5. Make a raw composition r_0 based on *s*;
6. Adjust for recurring formulas;
7. Adjust for singability;
8. Transcribe and perform.

These steps will be explained in detail in the following subsections.

2.1 Step 1: Represent the Neumes as Contour Letters

To represent the melodic contour information of the tenth-century neumatic notation, we designed a representation with an alphabet of six symbols, {h, l, e, o, b, p}, each representing a note of the (lost) melody. This representation can be considered an extension of the Parsons code (Parsons, 1975; Randel, 2001; Maessen, 2015). We use the letter h for a note higher than the preceding note, the l for a note lower than the preceding note and the e for notes of equal pitch. The letter o represents the first note of each chant, and also all notes for which the relative height with respect to the previous note is not determinable. In principle, this is the case for the first note of each neume. This would imply many o's in the transcriptions. This is undesirable since in order to represent as much contour information as possible in the transcription, the number of o's should be as low as possible. In many cases, however, the vertical position of a neume provides some indication of the pitch height with respect to the previous neume, enabling us to avoid the use of an o for the first note of the neume. To cover some remaining uncertainty, we also defined the letters b and p, respectively representing notes higher or equal, and lower or equal.

Apart from these letters, dashes and numerals are used for interpunction: 1 indicates the beginning of the chant, 5 the end, 4 the end of a main part of the melody and 3 a division within main parts. Three consecutive dashes, ---, indicate the beginning of a new word; two dashes, --, a new syllable and one dash, -, a new neumatic group.

We designate the lost melody represented by the tenth-century neumes with *x*. The transcription into contour representation we call *t*.

2.2 Step 2: Construct a Database with Pitched Melodies

Since eleventh and twelfth-century manuscripts do not provide information about rhythm and meter, but only about pitch, embellishments and interpunction, we are able to use the encoding developed for the music font Volpiano[2] to represent the melodies of pitch-readable chants in the database. In Volpiano font the characters 8, 9, a, b, c till o, p, q, r, s, represent the pitches F, G, A, B, c' till g'', a'', b'', c''', d''' on a five-line staff, the i being the flat sign on the third line. Numerals and dashes represent the interpunction as described in Section 2.1. For example, if we typeset the string "1---h-j-k-" in Volpiano font, we obtain .

We designate the database with *C*, and the *i*-th chant in the database with c_i.

Currently *C* is a subset of all chants referred to in Table 1 (see Table 1). For each of the five traditions all chants of four specific genres (tracts, cantus, benedictiones and offertories) are included. These belong to the longest chants of these traditions. From each tradition, several other chants are included as well, making *C* a collection of nearly 800 chants, good for almost 250,000 notes.

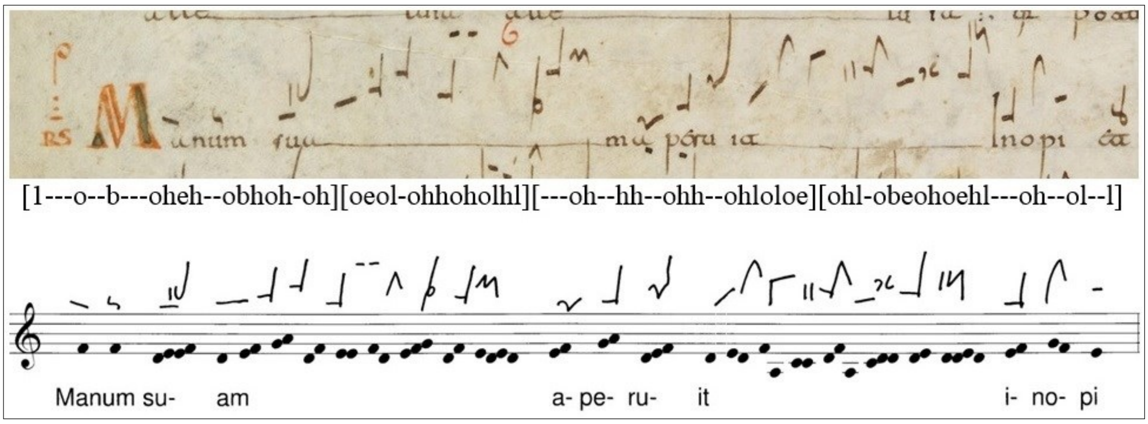

RS
Manum sua
mu pōru ia
Inopi ea
[1---o--b---oheh--obhoh-oh][oeol-ohhoholhl][---oh--hh--ohh--ohloloe][ohl-obeohoehl---oh--ol--l]
Manum su-
am
a- pe- ru-
it
i- no- pi

Figure 4. Beginning of the responsory *Manum suam aperuit* for St. Martin; top: the neumes of León, early X c.; between square brackets the transcription to contour letters; bottom: the final composition with León transcription.

[2] Volpiano font has been developed by David Hiley and Fabian Weber at the Institut für Musikwissenschaft of Regensburg University, it is downloadable from: http://www.uni-regensburg.de/Fakultaeten/phil_Fak_I/Musikwissenschaft/cantus/

2.3 Step 3: Divide *t* into Phrases

It is of course possible for the database to contain a melody c_i that fully corresponds with the contour t, and thus might be the lost melody x, but we consider this very unlikely. Nevertheless, our method is designed in such a way that *if* the lost melody is included in the database, we will find it. In any case, we do expect to be able to find a c_i that provides melodic material that sufficiently corresponds with t. In most cases, this will not be a correspondence between the complete contour t and c_i, but parts of c_i may correspond to parts of t. Therefore, our strategy is to manually divide t into melodic phrases. In our contour representation, these phrases are represented by square brackets (see Figure 4 for an example). We designate the j-th phrase in t with t_j.

To perform the division into phrases, we need an indication of the optimal phrase length. Phrases of one note would match to every melody of equal or greater length. Phrases of the full length of t would, in most cases, match no melody in C at all. Experimentally we found the best results for phrases between 9 and 18 notes, with the optimum around 12. We perform the segmentation into phrases by hand, as much as possible in accordance with the grammatical and musical syntax of x, as witnessed by the early neumatic notation. For different purposes the length of the phrases may be different (see Section 2.5).

2.4 Step 4: Choose a Source Melody

In our method, the final composition r (result) is based on a melody c_i that is closest to x: the source melody s.

Given the specific features of the transcribed chant t, and knowing the repertoire, we could choose a source melody c_i by hand from similar chants in related pitch-readable traditions GRE, MIL, ROM, BEN and MOZ (see Table 1). The similarity would be based on liturgical assignment, text, genre, and structure. Sometimes we may also have information about mode or historic relations of the lost chant, in which case we could prefer a source chant in a specific mode from a specific tradition.

Instead of choosing a single source melody by hand, however, it is preferable to create a database C' based on all these features – where C' is a subset of C – and automatically retrieve a suitable source melody s from C'. To this end, we computationally search matches for all phrases t_j of t in the melodies of C'. We implemented a brute force string matching algorithm to perform this search. The algorithm finds exact matches for a given t_j by comparing the sequence of contour letters in t_j with the notes of the melody string c_i at all possible positions in c_i. To compare a contour letter in t_j with a note in c_i, the interval that the note in c_i makes with a previous note is used. In case of a skip (explained below), not the direct preceeding note, but a more previous note is used. As a consequence, we cannot use a representation of the melody in which each note is represented as the interval with the direct preceeding note. This also prevents us from using a standard alignment algorithm on sequences of intervals.

The algorithm allows to skip a maximum of n_{skip} notes of c_i between each consecutive pair of matching notes in c_i. n_{skip} is a user-provided parameter of the algorithm. The higher n_{skip}, the more likely we find a match for a phrase t_j. However, for higher n_{skip}, the execution time increases, as well as the possibility to obtain perceptually unexpected successions of notes (see Section 2.7). For larger databases, n_{skip}=0 or n_{skip}=1 should be preferred. By using an exhaustive search rather than, for example, a string alignment approach, we are able to exactly control the maximum number of skipped notes.

The aim is to find the melody c_i that has matches for as many phrases t_j as possible, while each of these matches is as good as possible. To assess the quality of a phrase match, we introduce a numerical score based on two properties of the match: the total number of skipped notes and the position of the match within the full melody.

Figure 5 shows an example of a match m_j for contour `[olhollhohoh]` in a database melody c_i. Two notes of c_i needed to be skipped for the contour t_j to match the sequence of pitches.

For each phrase t_j in t, we compute the scores S_{skip} and S_{pos}. The score concerning skips is computed as

$$S_{skip} = max\left(0, 1 - \frac{skips}{\frac{1}{2}\left|t_j\right|}\right),$$

where *skips* is the total number of skipped notes, and $\left|t_j\right|$ is the length of phrase t_j from t. We include a factor ½ because we consider it undesirable to have more skips than half of the length of the phrase. In the example in Figure

c_i	…	k	h	l	h	k	j	h	h	g	k	j	h	k	l	m	m	l	…
m_j				l	h	k	j	h		g	k	j		k	l	m			
t_j			[	o	l	h	o	l		l	h	o		h	o	h	]		

Figure 5. Example of a match m_j for phrase t_j in database melody c_i. Two skips were needed to fit contour t_j to the notes of c_i.

5, the total number of skips is 2 and the length of t_j is 11, so the score S_{skip} is 0.64.

The score for position, S_{pos}, is computed as:

$$S_{pos} = 1 - \left|pos(t_j) - pos(m_j)\right| ,$$

where $pos(t_j)$ is the position of the first note of t_j relative to the full length of t, and $pos(m_j)$ is the position of the first note of m_j relative to the full length of c_i. $pos(t_j)$ and $pos(m_j)$ both are real numbers in [0,1], where 0 is the position of the first note and 1 the position of the last note.

We compute one score for the entire melody c_i according to

$$S = \lambda\overline{S_{pos}} + (1-\lambda)\overline{S_{skıp}} ,$$

where $\overline{S_{pos}}$ and $\overline{S_{skıp}}$ are averages over the phrases, and λ is a regularization parameter that determines the relative weight of the scores. Experimentally, we found $\lambda = 0.7$ to be a good value.

A phrase t_j might have more than one match in a database melody c_i. In that case, we obtain multiple scores for c_i, one for each possible configuration of matching phrases.

The scoring scheme is designed such that if the lost melody x is present in the database, we will find it, since in that case we will not need any skips, and the positions of the phrases will exactly correspond.

The melody s, which will serve as the source melody, is the melody in the database that obtained the highest score S.

2.5 Step 5: Make a Raw Composition

Given the source melody s found in step 4, we make a raw composition r_0 for t. This is the sequence of closest matches in s for the sequence of phrases in t. The procedure to construct r_0 is completely analogue to the procedure to find s in step 4. The crucial difference is that we replace the database C with the single source melody s, such that we find matches for all phrases of t in s. In the case that we do not find a match for a phrase t_j, we increase n_{skip}, accepting longer skips. If necessary, we also can adapt the segmentation of t to obtain shorter or longer phrases. By following this procedure, we get a composition for the lost melody that is entirely based on the melodic material from one other historic melody.

2.6 Step 6: Adjust Formulas

Many of the chants in neumatic notations will exhibit recurring patterns, *formulas* (intra-opus repeated segments) of between approximately 7 to 18 notes. As is the case in the five pitched traditions, it may be preferable to give specific formulas in all (or at least most) instances the same melodic content in our composition. To do so, we proceed as follows: we detect the formulas manually, we either manually pick a preferred pitch sequence (from r_0), or we calculate the closest matching sequence (as in step 4) of t_j in s, where t_j is a transcription of the formula to contour letters, and substitute this for the preferred formulas. The resulting melody we designate with r_1.

2.7 Step 7: Correct for Singability

Although steps 5 and 6 may seem to include some arbitrary decisions, the most subjective step is 7. It consists in singing r_1 and deciding (if necessary) to adjust the melody in minor details, or, in some cases, even greater parts. For many melodies this is necessary, since some pitches of r_1 may be judged to be very uncharacteristic for the style, especially at beginning and end of phrases. It may be good to "normalize" these manually, in agreement with our knowledge of medieval style. For the same reason, it may also be good to transpose some phrases a second, a fourth or a fifth. The criterion for these adjustments, however, should be that a small change should generate a great improvement in the melody, thus producing the final composition r. The number of adjustments can be considered an indication for the quality of our composition. We judge adjustments above 5 % of the total number of notes to become problematic. For those cases, we better start new calculations with different phrase divisions, other source chants and/or other pitch sequences for formulas. The 36 chants in Maessen (2016) have an average correction percentage of 3.95 for only first calculations. For the complete melody of Figure 4 we needed only 1 correction in 195 notes (we lowered the final pitch from e to d).

2.8 Step 8: Transcribe and Perform

After completion of the final composition r the musical score must be produced to enable others to sing the chant. Since our database consists of chants represented in Volpiano font (see Section 2.2), our compositions are also in Volpiano font. Therefore, we can easily transform our compositions to printable scores. Copying the original neumes above the score (see Figure 4) has proven to be a good way to give directors and singers inspiration for the manner of performance, especially concerning duration of notes and embellishments. The influential semiological interpretation of Gregorian neumatic notation (Cardine, 1968; González-Barrionuevo, 2015) can be considered important for their interpretation. See the original neumatic notation running along with the performance of Example 4 (and some other chants) by Gregoriana Amsterdam, uploaded to YouTube.[3]

3. EVALUATION

We do not claim lost melodies of the Mozarabic rite to survive in any number in pitch notation. We neither claim

[3] Accessible at: https://www.youtube.com/lelalilu

they did not. All steps in our method are developed with the possibility in mind that *if* the lost melody x would be included in our database, our final composition r will represent it. Maessen (2015) gives some examples of the original melody x found this way. Melodies very close to x may also be represented by r. However, although our method always finds a melody, not all possibly related melodies to x will be found. Many closely related melodies simply disagree too much to be found with our method. For the melodies we do find, several indications can be a sign that we are on the wrong track for recovering the lost melody. The more we have to shorten our phrases, increase n_{skip}, or have to adjust notes to singability, the less likely our composition is related to the lost melody.

In order to make any claim on the relationship with the lost chant, we could have developed a rating system in which all these kinds of aspects for each composition were combined. We chose, however, not to proceed this way, because our main concern is not any authenticity claim. Our concern is the semi-automatic production of singable melodies agreeing in all detail with our knowledge of the early notation. We believe that we can only understand something of the deeper layers of the lost tradition through the singing of its chant. Even without any authenticity claim we can experience many aspects of the lost musical tradition through its singing. Notably, the way the texts interact with the alternation of syllabic and melismatic passages, and the way recurring formulas interact with non-formulaic passages.

Since our first compositions in 2014, we are working on the improvement of the method in order to come to better melodic results. This paper presents the state of affairs in February 2017. An apparent weakness in our method still is the fact we do not use the interpunction encoded in Section 2.1. This information could make it possible to align not only single notes, but also, syllables, words and even sentences. There are many other problems we are still working on. We are also experimenting with generative probabilistic models and pattern detection algorithms. We do think, however, that our method, even in this stage, might be relevant for the comparison of other oral musical genres where rhythm and meter are not clearly prescribed in notation. Apart from Western and Eastern medieval liturgical music, we can think e.g. about troubadour songs.

Although it is not our main focus, and there is still a lot of work to be done, we can get some idea of the authenticity of our compositions when we consider the score S we obtain in step 4 (see Section 2.4). A score of 1 would indicate a 100 % agreement of the phrases of contour transcription t with our source melody s, i.e., no skips would be needed and the positions of the matches in s would exactly agree to the positions of the phrases in t. In most cases (average and more complex chants) this would (most likely) mean that we would have found the lost melody x. Under specific conditions, scores greater than 0.7 could indicate a close resemblance to the lost melody.[4] Up till now we recomposed and performed over 100 Mozarabic chants of diverse nature for several occasions. Most scores we found are below 0.4, indicating that we did not find lost melodies. Our compositions, however, can be sung. Some people even think some of them are beautiful. In all cases, however, our compositions proof perfectly suitable for liturgical practice. Examples of complete compositions can be found on *YouTube* and in Maessen (2015 & 2016).

4. REFERENCES

Cardine, E. (1968). Semiologia Gregoriana. Rome: Pontificium Institutum Musicae Sacrae. Translated as: *Gregorian Semiology*. Solesmes: Abbaye Saint-Pierre de Solesmes.

Driessen, M. (2013). Canticum Trium Puerorum, for 31-tone Fokker-Huygens organ, commissioned by Stichting Huygens-Fokker. Premièred on Feb. 10th 2013 at Muziekgebouw aan 't IJ - Amsterdam by Ere Lievonen.

Fernández de la Cuesta, I., Álvarez Martínez, R., & Llorens Martín, A. (2013). El canto mozárabe y su entorno. Estudios sobre la música de la liturgia viejo hispánica. Madrid: Sociedad Española de Musicología.

González-Barrionuevo, H. (2015). The Simple Neumes of the León Antiphonary. *Calculemus et Cantemus, Towards a Reconstruction of Mozarabic Chant*. Amsterdam: Gregoriana Amsterdam (pp. 31-52).

Hiley, D. (1993). Western Plainchant, A Handbook. Oxford: Clarendon Press.

Levy, K. (1998). Toledo, Rome, and the Legacy of Gaul. *Gregorian Chant and the Carolingians*. Princeton: Princeton University Press (pp. 31-81).

Maessen, G. (2015). Calculemus et Cantemus, Towards a Reconstruction of Mozarabic Chant. Amsterdam: Gregoriana Amsterdam.

Maessen, G. (2016). Saint Martin de Tours, Office mozarabe et Messe gallicane, Une reconstruction computationelle des chants. Amsterdam: Gregoriana Amsterdam.

Parsons, D. (1975). The Directory of Tunes and Musical Themes. Cambridge: Spencer Brown.

Randel, D. (2001). Mozarabic Chant, *The New Grove Dictionary of Music and Musicians (second edition)*. http://www.oxfordmusiconline.com/subscriber/article/grove/music/19269.

Rojo, C. & Prado, G. (1929). El Canto Mozárabe, Estudio histórico-critico de su antigüedad y estado actual. Barcelona: Diputación Provincial de Barcelona.

Swaan, E. (2012). Melodieën, lijnen, neumen en het probleem van het interpreteren van oude kerkgezangen. *Wim van Gerven, pionier van de gregoriaanse semiologie*. Amsterdam: Gregoriana Amsterdam (pp. 48-54).

[4] Maessen (2016) illustrates these conditions.

SYNTHESIS OF TURKISH MAKAM MUSIC SCORES USING AN ADAPTIVE TUNING APPROACH

Hasan Sercan Atlı, Sertan Şentürk
Music Technology Group
Universitat Pompeu Fabra
{hasansercan.atli, sertan.senturk}
@upf.edu

Barış Bozkurt
University of Crete
barisbozkurt0
@gmail.edu

Xavier Serra
Music Technology Group
Universitat Pompeu Fabra
xavier.serra
@upf.edu

ABSTRACT

Music synthesis is one of the most essential features of music notation software and applications aimed at navigating digital music score libraries. Currently, the majority of music synthesis tools are designed for Eurogenetic musics, and they are not able to address the culture-specific aspects (such as tuning, intonation and timbre) of many music cultures. In this paper, we focus on the tuning dimension in musical score playback for Turkish Makam Music (TMM). Based on existing computational tuning analysis methodologies, we propose an automatic synthesis methodology, which allows the user to listen to a music score synthesized according to the tuning extracted from an audio recording. As a proof-of-concept, we also present a desktop application, which allows the users to listen to playback of TMM music scores according to the theoretical temperament or a user specified reference recording. The playback of the synthesis using the tuning extracted from the recordings may provide a better user experience, and it may be used to assist music education, enhance music score editors and complement research in computational musicology.

1. INTRODUCTION

A music score is a symbolic representation of a piece of music that, apart from the note symbols, it contains other information that helps put those symbols into proper context. If the score is machine-readable, i.e. the elements can be interpreted by a music notation software, the different musical elements can be edited and sonified. This sonification can be done using a synthesis engine and with it, the users get an approximate real-time aural feedback on how the notated music would sound like if played by a performer.

Currently, most of the music score synthesis tools render the audio devoid of the performance added expression. It can be argued that this process provides an exemplary rendering reflecting theoretical information. However, the music scores of many music cultures do not explicitly include important information related to performance aspects such as the timing, dynamics, tuning and temperament. These characteristics are typically added by the performer, by using his or her knowledge of the music, in the context of the performance. Some aspects of the performance, such as the tuning and temperament, may differ due to the musical style, melodic context and aesthetic concerns. In performance-driven music styles and cultures, the "theoretical“ rendering of a music score might be considered as insufficient or flawed.

In parallel, the mainstream notation editors are currently designed for Eurogenetic musics. While these editors provide a means to compose and edit music scores in Western notation (and sometimes in other common notation formats such as tablatures), the synthesis solutions they provide are typically designed for 12 tone-equal-tempered (TET) tuning system, and they have limited support to render intermediate tones and mictotonal intervals. The wide use of these technologies may negatively impact the music creation process by introducing a standardized interpretation and it might even lead to loss of some variations in the expression and understanding of the music culture in the long term (McPhail, 1981; Bozkurt, 2012).

For such cases, culture-specific information inferred from music performances may significantly improve music score synthesis by incorporating the flexibility inherent in interpretation. In this study, we focus on the tuning and temperament dimensions in music score synthesis, specifically for the case of Turkish makam music (TMM). Turkish makam music is a suitable example since performances use diverse tunings and microtonal intervals, which vary with respect to the makam (melodic structure), geographical region and artists. Based on an existing computational tuning analysis methodology, we propose an adaptive synthesis method, which allows the user to synthesize the melody in a music score either according to a given tuning system or according to the tuning extracted from audio recordings. In addition, we have developed a proof-of-concept desktop application for the navigation and playback of the music scores of TMM, which uses the adaptive synthesis method we propose. To the best of our knowledge, this paper presents the first work on performance-driven synthesis and playback of TMM.

For reproducibility purposes, all relevant materials such as musical examples, data and software are open and publicly available via the companion page of the paper hosted in the Compmusic Website. [1]

The rest of the paper is structured as follows: Section 2 gives a brief information of TMM. Section 3 presents an overview of the relevant commercial music synthesis software and the academic studies. Section 4 explains the

[1] http://compmusic.upf.edu/node/339

ISBN: 978-84-697-2303-6

methodology that adapts the frequencies of the notes in a machine readable music score to be synthesized and the preparation of the tuning presets. Section 5 explains the music score collection, the implementation of the methodology and the desktop software developed for discovering the score collection. Section 6 wraps up the paper with a brief discussion and conclusion.

2. TURKISH MAKAM MUSIC

Most of the melodic aspects of TMM can be explained by the term *makam*. Each makam has a particular scale, which gives the "lifeless" skeleton of the makam (Signell, 1986). Makams are modal structures (Powers, et al., 2013), which gains its character through its melodic progression (*seyir* in Turkish) (Tanrıkorur, 2011). Within the progression, the melodies typically revolve around an initial tone (*başlangıç* or *güçlü* in Turkish) and a final tone (*karar* in Turkish) (Ederer, 2011; Bozkurt et al., 2014).

Karar is typically used synonymous to *tonic*, and the performance of a makam ends almost always on this note. There is no definite reference frequency (e.g. $A4 = 440$Hz) to tune the performance tonic. Musicians might choose to perform the music in a number of different transpositions (*ahenk* in Turkish), any of which might be favored over others due to instrument/vocal range or aesthetic concerns (Ederer, 2011).

There are several theories attempting to explain the makam practice (Arel, 1968; Karadeniz, 1984; Özkan, 2006; Yarman, 2008). Among these, Arel-Ezgi-Uzdilek (AEU) theory (Arel, 1968) is the mainstream theory. AEU theory is based on Pythagorean tuning (Tura, 1988). It also presents an approximation for intervals by the use of Holderian comma (Hc)[2] (Ederer, 2011), which simplifies the theory via use of discrete intervals instead of frequency ratios. "Comma" (*koma* in Turkish) is part of daily lexicon of musicians and often used in education to specify intervals in makam scales. Some basic intervals used in AEU theory are listed in Table 1 (with sizes specified in commas on the last column)

Since early 20th century, a score representation extending the traditional Western music notation has been used as a complement to the oral practice (Popescu-Judetz, 1996). The extended Western notation typically follows the rules of AEU theory. Table 2 lists the accidental symbols specific to TMM used in this notation.

The music scores tend to notate simple melodic lines and the musicians follow the scores of the compositions as a reference. Nevertheless, they extend the notated "musical idea" considerably during the performance by adding non-notated embellishments, inserting/repeating/omitting notes, altering timing, and changing the tuning and temperament. The temperament of some intervals in a performance might differ from the theoretical (AEU) intervals as much as a semi-tone (Signell, 1986).

Name	Flat	Sharp	Hc
Koma	𝄳	𝄲	1
Bakiye	♭	♯	4
Küçük mücennep	♭	♯	5
Büyük mücennep	♭	♯	8

Table 1: The accidental symbols defined in extended Western notation used in TMM, their theoretical intervals in Hc according to the AEU theory.

3. BACKGROUND

Many commercial music notation software tools such as Sibelius[3], Finale[4] and MuseScore[5] support engraving and editing the accidentals used in Turkish makam music. However, they provide no straightforward or out-of-the-box solution for microtonal synthesis. For example, MuseScore only supports synthesis of 24 tone-equal-temperament system, which is not sufficient to represent the the intervals in either TMM practice or theory.

Mus2[6] is a music notation software specifically designed for the compositions including microtonal content. It includes a synthesis tool that allows users to playback music scores in different microtonal tuning systems such as just intonation. In addition, Mus2 allows the users to modify the intervals manually. Nevertheless, manually specifying the intervals could be tedious. In addition, the process may not be straightforward for many users, which do not have a sufficient musical, theoretical or mathematical background.

There exists several studies in literature for automatic tuning analysis of TMM (Bozkurt, 2008; Gedik & Bozkurt, 2010) and Indian art musics (Serrà et al., 2011; Koduri et al., 2014). These studies are mainly based on pitch histogram analysis. Bozkurt et al. (2009) analyzed the recordings of masters in 9 commonly performed makams by computing a pitch histogram from each recording and then detecting the peaks of histograms. Considering each peak as one of the performed scale degrees, they compared the identified scale degrees with the theoretical ones defined in several theoretical frameworks. The comparison showed that the current music theories are not able to explain the intervallic relations observed in the performance practice well.

Later, Bozkurt (2012) proposed an automatic tuner for TMM. In the tuner, the user can specify the makam and input an audio recording in the same makam. Then, the tuning is extracted from the audio recording using the pitch histogram analysis method described above. The tuning information is then provided the user interactively, while she/he is tuning an instrument. Similarly, Şentürk et al. (2012) has incorporated the same pitch histogram based tuning analysis methodology into an audio-score alignment methodology proposed for TMM. In this method, the tuning of the audio recording is extracted as a preprocessing

[2] i.e. 1 Hc = $\frac{1200}{53} \approx 22.64$ cents

[3] http://www.avid.com/sibelius

[4] http://www.finalemusic.com/

[5] https://musescore.org/

[6] https://www.mus2.com.tr/en/

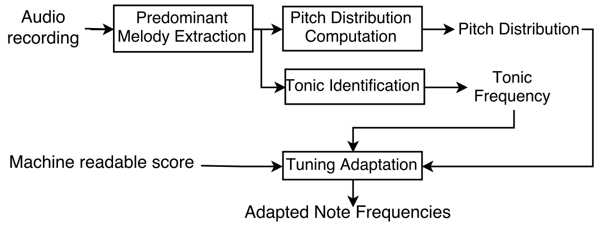


Figure 1: The flow diagram of the adaptive tuning methodology

step prior to the alignment step. Next, it is used (instead of the theoretical temperament) to generate a synthetic pitch track from the relevant music score. This step minimizes the temperament differences between the audio predominant melody and the synthetic pitch track, and therefore a smaller cost is emitted by the alignment process.

4. METHODOLOGY

The proposed system differs from existing synthesizers by allowing the user to supply a reference recording (for temperament) from which the intervals may be learned automatically. When a reference recording is not available, our method maps the note symbols according to the intervals described in the music theory with respect to the user provided tonic frequency. If the user provides a reference audio recording, our method first extracts the predominant melody from the audio recording. Next, it computes a pitch distribution from the predominant melody and identifies the frequency of tonic note in the performance. By applying peak detection to the pitch distribution, our method obtains the stable frequencies performed in the audio recording. Then, the stable pitches are mapped to the note symbols in the music score by taking the identified tonic frequency as the reference. Finally, synthesis is performed by using the Karplus-Strong string synthesis method. The flow diagram of the adaptive tuning method is shown in Figure 1.[7]

4.1 Predominant Melody Extraction

To identify the tuning, the method first extracts the predominant melody of the given audio recording. We use the methodology proposed in (Atlı et al., 2014).[8] It is a variant of the methodology proposed in (Salamon & Gómez, 2012), which is optimized for TMM. Then, we apply a post-filter[9] proposed in (Bozkurt, 2008) on the estimated predominant melody. The filter corrects the octave errors. It also removes the noisy regions, short pitch chunks and extreme valued pitch estimations of the extracted predominant melody.

4.2 Pitch Distribution Computation

Next, we compute a pitch distribution (PD) (Chordia & Şentürk, 2013) from the extracted predominant melodies (Figure 2). The PD shows the relative occurrence of the frequencies in the extracted predominant melody.[10] We use the parameters described for pitch distribution extraction in (Şentürk, 2016, Section 5.5) as follows: The bin size of the distribution is set as 7.5 cents $\approx 1/3$ Hc resulting in a resolution of 160 bins per octave (Bozkurt, 2008). We use kernel density estimation and select the kernel as normal distribution with a standard deviation of 7.5. The width of the kernel is selected as 5 standard deviations peak-to-tail (where the normal distribution is greatly diminished) to reduce computational complexity.

4.3 Tonic Identification

In parallel, we identify the tonic frequency of the performance using the methodology proposed by Atlı et al. (2015). The method identifies the frequency of the last performed note, which is almost always the tonic of the performance (Section 2). The method is reported to give highly accurate results, i.e. $\sim 89\%$ in (Atlı et al., 2015).

4.4 Tuning Analysis and Adaptation

We detect the peaks in the PD using the peak detection method explained in (Smith III & Serra, 1987).[11] The peaks could be considered as the set of stable pitches performed in the audio recording (Bozkurt et al., 2009). The stable pitches are converted to scale degrees in cent scale by taking the identified tonic frequency as the reference using the formula:

$$c_i = 1200 \log_2(f_i/t) \qquad (1)$$

where f_i is the frequency of a stable pitch in Hz, t is the identified tonic frequency and c_i is the scale degree of the stable pitch in cents.

In parallel, the note symbols in the scale of the makam[12] is inferred from the key signature of the makam[13] and extended to ± two octaves. The note symbols are initially mapped to the theoretical temperaments (scale degrees in cents) according to the AEU theory (e.g. if the tonic symbol is G4, the scale degree of A4 is 9 Hc ≈ 203.8 cents).

Next, the performed scale degrees are matched with the theoretical scale degrees using a threshold of 50 cents (close to 2.5 Hc, which is reported as an optimal by (Bozkurt et al., 2009)). If a performed scale degree is close to more than one theoretical scale degree (or vice versa), we only match the closest pair. If there are no matches for a theoretical scale degree, we keep the theoretical value. As a trivial addition to (Bozkurt et al., 2009), we re-map the

[7] The implementation of our methodology is openly available at https://github.com/hsercanatli/symbtrsynthesis.

[8] The implementation is available at https://github.com/sertansenturk/predominantmelodymakam.

[9] The implementation is available at https://github.com/hsercanatli/pitchfilter.

[10] We use the implementation presented in (Karakurt et al., 2016): https://github.com/altugkarakurt/morty.

[11] The implementation is available in Essentia (Bogdanov et al., 2013): http://essentia.upf.edu/.

[12] The makam is known from the music score (Section 4.5).

[13] Available at https://github.com/miracatici/notemodel/blob/v1.2.1/notemodel/data/makam_extended.json.

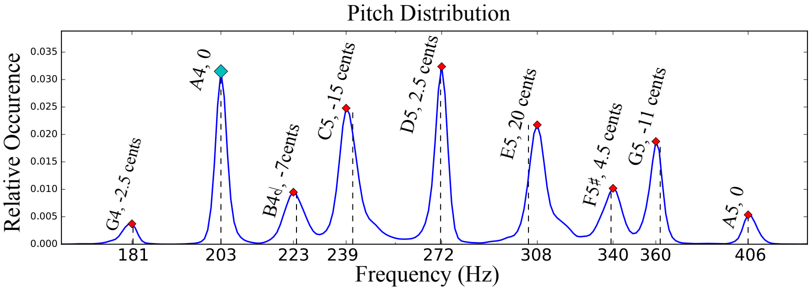


Figure 2: The tuning extracted from a recording in Hüseyni makam, performed by Tanburi Cemil Bey.

theoretical scale degrees to the note symbols and obtain the < *note symbol - stable pitch* > pairs. [14]

Figure 2 [15] shows an example tuning analysis applied on a historical recording in Hüseyni makam performed by Tanburi Cemil Bey. [16] The frequency of each stable note is shown on the x-axis. The vertical dashed lines indicate the frequencies of the notes according to the theoretical intervals. The matched note symbol and the deviation from the theoretical scale degree of each stable pitch is displayed right next to the corresponding peak on the PD. It can be observed that the some of the notes - esp. çargah (C5) and hüseyni (E5) notes - substantially deviate from the AEU theory.

4.5 Score Synthesis

From the machine-readable music score, we read the note sequence, nominal tempo, makam and tonic symbol (last note in the sequence). The note symbols are converted to a stable pitches, by referring to the < *note symbol - stable pitch* > pairs obtained from tuning analysis. In parallel, the symbolic note durations are converted to seconds by referring to the nominal tempo. Next, we generate a pitch-track from the note sequence in the score by sampling the mapped stable pitches relative to the their duration in seconds at a frame rate of 44100 Hz and then concatenating all samples (Şentürk et al., 2012). The score pitch-track is synthesized using the Karplus-Strong string synthesis (Jaffe & Smith, 1983). [17]

In addition, we mark the sample index of each note onset in the score pitch-track to later use to synchronize the music score visualization during playback in our desktop application (Section 5.4).

5. APPLICATION

As a proof-of-concept, we have developed a desktop application [18] , for the navigation and playback of the music scores of TMM. In this section, we showcase the application (Section 5.4) and discuss how it fits into the Dunya ecosystem, which comprises all the music corpora and related software tools that have been developed as part of the CompMusic project. In specific, we describe the music score collection (Section 5.1), the tuning presents extracted from audio recordings (Section 5.2), and the data processing and storage platform hosted on web (Section 5.3).

5.1 Music Scores

In this study, we use the music scores in the SymbTr score collection (Karaosmanoğlu, 2012). [19] SymbTr is currently the most representative open-source machine-readable music score collection of TMM (Uyar et al., 2014). Specifically, we use the scores in MusicXML format. This format is preferred, because it is commonly used in many music notation and engraving software. The scores in MusicXML format does not only contain the notes, but also other relevant information such as the sections, tempo, composer, *makam* and *form* of the related musical piece. We use some of this information to search the scores in the desktop application (Section 5.4).

To render the music scores during playback (Section 5.4), we first convert the scores in MusicXML format to LilyPond and then to SVGs. [20] Each note element in the SVG score contains the note indices in the MusicXML score.

5.2 Tuning Presets

Using the methodology described in Section 4, we extracted the tuning from 10 "good-quality" recordings as presets for each of the Hicaz, Nihavent, Uşşak, Rast and Hüzzam makams (i.e. 50 recordings in total). [21] These are the most commonly represented makams in the SymbTr collection, and they constitute more than 25% of the music scores in the SymbTr collection (Şentürk, 2016, Table 3.2). The recordings are selected from the CompMusic Turkish makam music audio collection (Uyar et al., 2014), which is

[14] The implementation is available at `https://github.com/miracatici/notemodel`.

[15] The Figure and the explanation is reproduced from (Şentürk, 2016, Section 5.9).

[16] `https://musicbrainz.org/recording/8b8d697b-cad9-446e-ad19-5e85a36aa253`

[17] We modified the implementation of the Karplus-Strong model in the PySynth library: `https://github.com/mdoege/PySynth`.

[18] `https://github.com/MTG/dunya-desktop/tree/adaptive-synthesis`

[19] The SymbTr collection is openly available online: `https://github.com/MTG/SymbTr`.

[20] The score conversion code is openly available at `https://github.com/sertansenturk/tomato/blob/v0.9.1/tomato/symbolic/scoreconverter.py`.

[21] The recording metadata and the relevant features are stored in GitHub for reproducibility purposes: `https://github.com/MTG/otmm_tuning_intonation_dataset/tree/atli2017synthesis`.

currently the most representative audio collection of TMM, available for computational research.

We have synthesized 1222 scores according to the presets (in total 12220 audio synthesis) and all the 2200 scores with the theoretical tuning.

5.3 Dunya and Dunya-web

Dunya is developed with Django framework to store the data and execute the analysis algorithms developed within the CompMusic Project. [22] The audio recordings, music scores and relevant metadata are stored in a PostgreSQL database. [23] Its possible to manage information about the stored data and submit analysis tasks on the data from the administration panel. The output of each analysis is also stored in the database. The data can be accessed from the Dunya REST API. We have also developed a Python wrapper, called pycompmusic, [24] around the API.

To showcase our technologies developed within the CompMusic project, we have created a web application for music discovery called Dunya-web (Porter et al., 2013). The application displays the resulting automatic analysis. Dunya-web has a separate organization for each music culture studied within the CompMusic project. [25]

5.4 Dunya-desktop

In addition to Dunya-web, we have been working on a desktop application for accessing and visualizing the corpora created in the scope of CompMusic project. The aim is developing a modular and customizable music discovery interface to increase the reusability of the CompMusic research results to researchers.

Dunya-desktop [26] is directly connected to the Dunya Framework. The user could query the corpora and download the relevant data to the local working environment such as music scores, audio recordings, extracted features (predominant melody, tonic, pitch distribution and etc.). The interface provides an ability to create sub-collections to the user. It also comes with some visualization and annotation tools for extracted features and music score that the user could create a customized tool for his/her research task.

Our software is developed in Python 2.7/3 using PyQt5 [27] library. This library allows us to use the Qt5 binaries in Python programming language. The developed software is compatible with Mac OSX and GNU/Linux distributions.

The software that we developed as a proof-of-concept is an extension and customization of Dunya-desktop. The flow diagram of the user interaction in the desktop application is shown in Figure 3. The application allows the user to search a specific score by filtering metadata. If the selected composition is in one of the *makam*s with a preset, the user can choose to playback the score synthesized according to the AEU theory or to the available tuning presets. Otherwise, only the synthesis according to the AEU theory is available.

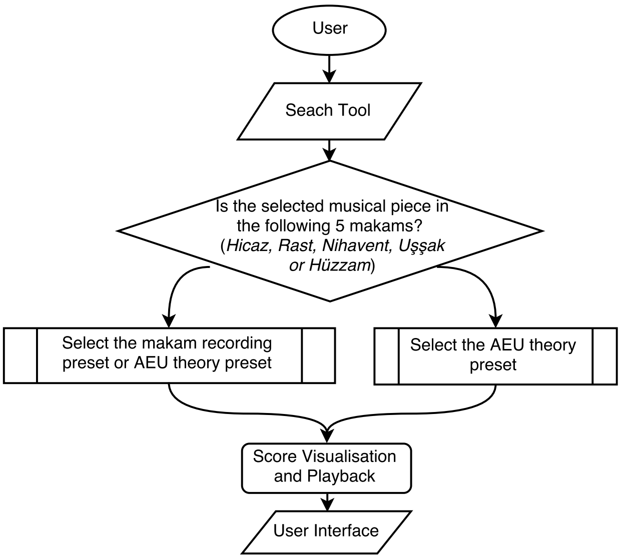

User
Seach Tool
Is the selected musical piece in the following 5 makams? (*Hicaz, Rast, Nihavent, Uşşak or Hüzzam*)
Select the makam recording preset or AEU theory preset
Select the AEU theory preset
Score Visualisation and Playback
User Interface

Figure 3: The flow diagram of the desktop software

A screenshot of the score playback window is shown in Figure 4. Remember that, we have the mapping between the synthesized audio and the note indices in the MusicXML score (Section 4.5) and also the mapping between the note indices in the MusicXML score and the SVG score (Section 5.1). Therefore, we can synchronize the SVG score and the synthesized audio. The current note in playback is highlighted in red on the rendered SVG score.

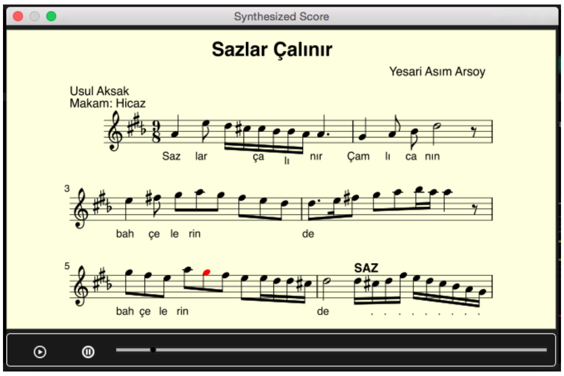


Figure 4: A screenshot of the playback window of the software

6. DISCUSSIONS AND CONCLUSIONS

In this paper, an automatic synthesis and playback methodology that allows users to listen a music score according to a given tuning system or according to the tuning extracted from a set of audio recordings is presented. We have also developed a desktop software that allows users to discover a TMM score collection. As a proof-of-concept,

[22] `https://www.djangoproject.com/`
[23] `https://www.postgresql.org/`
[24] `https://github.com/MTG/pycompmusic`
[25] for TMM: `http://dunya.compmusic.upf.edu/makam/`.
[26] `https://github.com/MTG/dunya-desktop`
[27] `https://wiki.python.org/moin/PyQt`

we apply the software on the SymbTr score collection. According to the feedback we have received from musicians and musicologists, the playback using the extracted tuning from a performance provides a better experience. In the future, we would like to verify this feedback quantitatively by conducting user studies. We would also like to improve the synthesis methodology by incorporating the score-informed tuning and intonation analysis (Şentürk, 2016, Section 6.11) obtained from audio-score alignment (Şentürk et al., 2014).

7. ACKNOWLEDGEMENTS

We would like to thank Burak Uyar for his contributions in converting the SymbTr scores from the original tabular format to MusicXML and Andrés Ferraro for his support in Dunya-web. This work is partially supported by the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

8. REFERENCES

Arel, H. S. (1968). *Türk Musikisi Nazariyatı*. ITMKD yayınları.

Atlı, H. S., Bozkurt, B., & Şentürk, S. (2015). A method for tonic frequency identification of Turkish makam music recordings. In *5th International Workshop on Folk Music Analysis (FMA)*, Paris, France.

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 (ATMM)*, Ankara, Turkey.

Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J., & Serra, X. (2013). Essentia: An audio analysis library for music information retrieval. In *Proceedings of 14th International Society for Music Information Retrieval Conference (ISMIR 2013)*, (pp. 493–498)., Curitiba, Brazil.

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. (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., 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.

Chordia, P. & Şentürk, S. (2013). Joint recognition of raag and tonic in North Indian music. *Computer Music Journal*, *37*(3).

Ederer, E. B. (2011). *The Theory and Praxis of Makam in Classical Turkish Music 1910-2010*. PhD thesis, University of California, Santa Barbara.

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

Jaffe, D. A. & Smith, J. O. (1983). Extensions of the Karplus-Strong plucked-string algorithm. *Computer Music Journal*, *7*(2), 56–69.

Karadeniz, M. E. (1984). *Türk Musıkisinin Nazariye ve Esasları*, (pp. 159). İş Bankası Yayınları.

Karakurt, A., Şentürk, S., & Serra, X. (2016). MORTY: A toolbox for mode recognition and tonic identification. In *Proceedings of the 3rd International Digital Libraries for Musicology Workshop (DLfM 2016)*, (pp. 9–16)., New York, NY, USA.

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).

Koduri, G. K., Ishwar, V., Serrà, J., & Serra, X. (2014). Intonation analysis of rāgas in carnatic music. *Journal of New Music Research*, *43*, 72–93.

McPhail, T. L. (1981). *Electronic colonialism: The future of international broadcasting and communication*. Sage Publications.

Özkan, I. H. (2006). *Türk mûsikısi nazariyatı ve usûlleri: Kudüm velveleleri*. Ötüken Neşriyat.

Popescu-Judetz, E. (1996). *Meanings in Turkish Musical Culture*. Istanbul: Pan Yayıncılık.

Porter, A., Sordo, M., & Serra, X. (2013). Dunya: A system for browsing audio music collections exploiting cultural context. In *Proceedings of 14th International Society for Music Information Retrieval Conference (ISMIR)*, Curitiba, Brazil.

Powers, et al., H. S. (accessed April 5, 2013). Mode. Grove Music Online.

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. (2016). *Computational Analysis of Audio Recordings and Music Scores for the Description and Discovery of Ottoman-Turkish Makam Music*. PhD thesis, Universitat Pompeu Fabra, Barcelona.

Şentürk, S., Holzapfel, A., & Serra, X. (2012). An approach for linking score and audio recordings in makam music of turkey. In *2nd CompMusic Workshop*, Istanbul, Turkey.

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

Serrà, J., Koduri, G. K., Miron, M., & Serra, X. (2011). Assessing the tuning of sung Indian classical music. In *12th International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 263–268)., Miami, USA.

Signell, K. L. (1986). *Makam: Modal practice in Turkish art music*. Da Capo Press.

Smith III, J. O. & Serra, X. (1987). *PARSHL: an analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation*. CCRMA, Department of Music, Stanford University.

Tanrıkorur, C. (2011). *Osmanlı Dönemi Türk Musikisi*. Dergah Yayınları.

Tura, Y. (1988). *Türk Musıkisinin Meseleleri*. Pan Yayıncılık, Istanbul.

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.

Yarman, O. (2008). *79-tone tuning & theory for Turkish maqam music*. PhD thesis, İstanbul Teknik Üniversitesi Sosyal Bilimler Enstitüsü.

Oral session 3

Folk Music Analysis

MÁLAGA, Spain

14-16 June 2017

IMPROVED ONSET DETECTION FOR TRADITIONAL IRISH FLUTE RECORDINGS USING CONVOLUTIONAL NEURAL NETWORKS

Islah Ali-MacLachlan, Carl Southall, Maciej Tomczak, Jason Hockman
DMT Lab, Birmingham City University
islah.ali-maclachlan, carl.southall, maciej.tomczak, jason.hockman
@bcu.ac.uk

ABSTRACT

The usage of ornaments is key attribute that defines the style of a flute performances within the genre of Irish Traditional Music (ITM). Automated analysis of ornaments in ITM would allow for the musicological investigation of a player's style and would be a useful feature in the analysis of trends within large corpora of ITM music. As ornament onsets are short and subtle variations within an analysed signal, they are substantially more difficult to detect than longer notes. This paper addresses the topic of onset detection for notes, ornaments and breaths in ITM. We propose a new onset detection method based on a convolutional neural network (CNN) trained solely on flute recordings of ITM. The presented method is evaluated alongside a state-of-the-art generalised onset detection method using a corpus of 79 full-length solo flute recordings. The results demonstrate that the proposed system outperforms the generalised system over a range of musical patterns idiomatic of the genre.

1. INTRODUCTION

Figure 1: Player with Rudall and Rose eight-key simple system flute manufactured from cocus wood.

Irish Traditional Music (ITM) is a form of Folk music that developed alongside social dancing and has been an integral part of Irish culture for hundreds of years (Boullier, 1998). ITM consists of various subgenres and is played with a wide variety of traditional instrumentation, including melody instruments such as fiddles, bagpipes, tin whistles, accordions and flutes. Figure 1 presents an ITM performer with a wooden simple system flute.

Determining the stylistic differences between players is an important first step towards understanding how the music and culture associated with ITM has developed. Within traditional music, mastery is determined by technical and artistic ability demonstrated through individuality and variation in performances. Individual playing style is comprised of several features, including variations in melody, rhythmic phrasing, articulation, and ornamentation (McCullough, 1977; Hast & Scott, 2004; Keegan, 2010; Köküer et al., 2014).

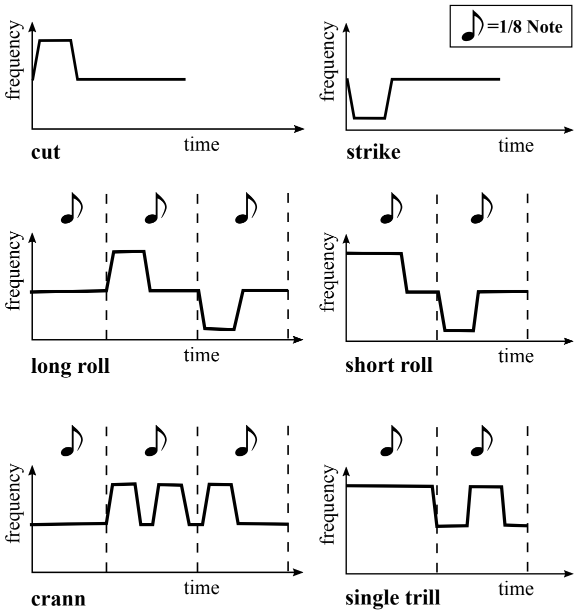


Figure 2: Frequency over time of *cut* and *strike* articulations showing change of pitch. *Long* and *short rolls*, *cranns* and *single trills* are also shown with pitch deviations. Eighth-note lengths are shown for reference.

Automated identification of a player's style would be useful in the musicological investigation of various trends within the ITM timeline. A first step towards automated style identification is the detection of onsets related to notes and ornaments. This study continues the work of Ali-MacLachlan et al. (2016) by evaluating notes and single-note ornaments known as *cuts* and *strikes*. We also investigate breaths and the cut and strike elements of multi-note ornaments known as *short roll*, *long roll*, *crann* and *single trill* as described in Larsen (2003). Figure 2 depicts single-note and multi-note ornaments over time.

Onset detection algorithms are used to identify the start of musically relevant events. Ornament onset detection for Irish traditional flute recordings is a difficult task due to their subtle nature; ornaments tend to be played in a short and soft manner, resulting in onsets characterised by a long attack with a slow energy rise (Gainza et al., 2005; Böck & Widmer, 2013).

1.1 Related work

There are relatively few studies concentrating on onset detection of flute signals within ITM. Gainza et al. (2004) and Kelleher et al. (2005) used instrument-optimised band-specific thresholds alongside a decision tree to determine note, cut or strike based on duration and pitch. Köküer et al. (2014) also analysed flute recordings, using an instrument-specific filterbank and a fundamental frequency estimation method using the YIN algorithm by De Cheveigné & Kawahara (2002) to minimise inaccuracies associated with octave doubling. More recently, Jančovič et al. (2015) presented a method for transcription of ITM flute recordings with ornamentation using hidden Markov models and Beauguitte et al. (2016) evaluated note tracking using a range of methods on a corpus of 30 tune recordings.

Onset detection techniques used in existing flute signal analysis have largely relied upon algorithms utilising signal processing, while state-of-the-art generalised onset detection methods use probabilistic modelling. Ali-MacLachlan et al. (2016) evaluated 11 methods that had previously performed well in the MIREX wind instrument class. `OnsetDetector` achieved the highest precision and F-measure scores. The use of bidirectional long short-term memory neural networks allows this model to learn the context of an onset based on past and future information, resulting in high performance in the context where soft onsets and features with small pitch deviations are coupled with other spurious events.

1.2 Motivation

The approach undertaken in this paper extends upon the work published in Ali-MacLachlan et al. (2016) in which onsets were detected through the use of the `OnsetDetector` system Eyben et al. (2010). Inter-onset segment classification was performed using an classification method based on a feed-forward neural network.

The `OnsetDetector` system was trained on a broad range of music making it effective at detecting a variety of instrument onsets. While note onset detection accuracy was very successful, ornament detection accuracies proved to be quite low by comparison. In an attempt to improve onset detection for ITM, we implemented an onset detection method based on a convolutional neural network (CNN) and trained this model specifically on ITM flute recordings. As we believe that the detection of ornament onsets to be context-dependent, we evaluate detection accuracy in relation to events that occur immediately before and after the detected events. This evaluation allows us to determine *where* onset detection errors occur and allows us to observe limitations in the detection of notes, cuts, strikes and breaths, in the context of traditional music being played authentically at a professional level.

The remainder of this paper is structured as follows: Section 2 outlines the proposed onset detection method and Section 3 presents our evaluation methodology and dataset. Section 4 presents the results of this evaluation and Section 5 presents conclusions and future work.

2. METHOD

Our onset detection method is based on a convolutional neural network (CNN) classification method. CNNs share weights by implementing the same function on sub-regions of the input. This enables CNNs to process a greater number of features at a lower computational requirement compared to other neural network architectures (i.e., multi-layer perceptron). High onset detection accuracies have been achieved by CNNs using larger input features (Schluter & Böck, 2014).

Figure 3 gives an overview of the implemented CNN architecture. The input features are first fed into two sets of convolutional and max pooling layers containing dropouts and batch normalisation. The output is then reshaped into a one-dimensional format before being run through a fully-connected layer and a softmax output layer.

2.1 Convolutional and max pooling layers

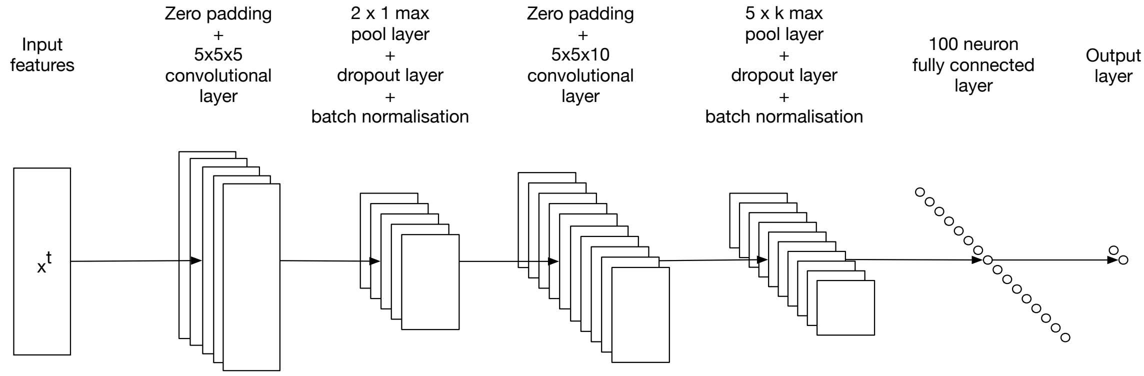
The output h of a two-dimensional convolutional layer with a rectified linear unit transfer function is calculated using:

$$h_{ij}^{f} = r\left(\sum_{l=0}^{L-1}\sum_{m=0}^{M-1} W_{ml}^{f} x_{(i+l)(j+m)} + b^{f}\right) \tag{1}$$

where x is the input features, W and b are the shared weights and bias and f is the feature map. L and M are the dimensions of the shared weight matrix and I and J are the output dimensions of that layer. The equation for the rectifier linear unit transfer function r is:

$$r(\phi) = max(0, \phi) \tag{2}$$

The output of the convolutional layer h was then processed using a max pooling layer which resulted in a $\frac{I}{a}$ by $\frac{J}{b}$ output where a and b are the dimensions of the sub-regions processed. A dropout layer (Srivastava et al., 2014) and batch normalisation (Ioffe & Szegedy, 2015) were then implemented.


Input features
Zero padding + 5x5x5 convolutional layer
2 x 1 max pool layer + dropout layer + batch normalisation
Zero padding + 5x5x10 convolutional layer
5 x k max pool layer + dropout layer + batch normalisation
100 neuron fully connected layer
Output layer
x^t

CNN5	512 x 5 x 1	512 x 5 x 5	256 x 5 x 5	256 x 5 x 10	51 x 5 x 10	100	2
CNN11	512 x 11 x 1	512 x 11 x 5	256 x 11 x 5	256 x 11 x 10	51 x 6 x 10	100	2
CNN21	512 x 21 x 1	512 x 21 x 5	256 x 21 x 5	256 x 21 x 10	51 x 7 x 10	100	2
CNN41	512 x 41 x 1	512 x 41 x 5	256 x 41 x 5	256 x 41 x 10	51 x 9 x 10	100	2
CNN101	512 x 101 x 1	512 x 101 x 5	256 x 101 x 5	256 x 101 x 10	51 x 11 x 10	100	2

Figure 3: Overview of the proposed implemented CNN system with different input feature sizes.

2.2 Fully-connected layer

A fully-connected layer consists of neurons which are linked to all of the neurons in previous and future layers. The output Y of a fully connected layer with a rectified linear unit transfer function is calculated using:

$$Y = r(W_c z + b_c) \tag{3}$$

where z is the input, W_c is the weight matrix and b_c is the bias. For the softmax output layer the rectified linear unit r transfer function is swapped for the softmax function which is calculated using:

$$softmax(\phi) = \frac{e^{\phi}}{\sum e^{\phi}} \tag{4}$$

2.3 Implementation

The CNN was implemented using the Tensorflow Python library (Abadi et al., 2016) with training data consisting of target activation functions created from ground truth annotations. A frame-based approach was taken where each frame is assigned 1 if it contains an onset or 0 if it does not.

2.4 Input features

Before processing by the CNN, the audio files must be segmented into frame-wise spectral features. An N sample length audio file was segmented into T frames using a Hanning window of γ samples ($\gamma = 1024$) and a hop size of $\frac{\gamma}{2}$. A frequency representation of each of the frames was then created using the discrete Fourier transform resulting in a $\frac{\gamma}{2}$ by T spectrogram. Various centred on the frame to be classified.

As classification is performed on the frame at the centre of the input features, a potentially crucial parameter is the number of input frames ψ. To determine the most efficient number of frames to use as the input for the CNN, five different values for ψ were used ($\psi = [5, 11, 21, 41, 101]$) creating the `CNN5`, `CNN11`, `CNN21`, `CNN41`, `CNN101` versions respectively.

2.5 Layer sizes

The layer sizes used for the different input features are indicated at the bottom of Figure 3. The size of all layers are consistent across systems apart from the second dimension k of the second max pooling layer. k is set to 1, 2, 3, 5 and 10 for the different input features sizes respectively.

2.6 Peak picking

The onsets must be temporally located from within the activation function Y output from the CNN. To calculate onset positions, the method from Southall et al. (2016) is used. A threshold τ is first determined using the mean across all frames and a constant λ:

$$\tau = \lambda \bar{Y} \tag{5}$$

The current frame t is determined to be an onset if its magnitude is greater than those of the surrounding two frames and above threshold τ.

$$O(t) = \begin{cases} 1, & y^t = max(y^{t-1:t+1}) \;\&\; y^t > \tau, \\ 0, & otherwise. \end{cases} \tag{6}$$

Finally, if an onset occurs within $25ms$ seconds of another then it is removed.

2.7 Training

The training data is divided into 1000 frame mini-batches consisting of a randomised combination of 100 frame re-

Player	Album(s)	Reels	Jigs	Polkas	Hornpipes
Harry Bradley	The First of May	8	4		4
Bernard Flaherty	Flute Players of Roscommon Vol.1	2			
John Kelly	Flute Players of Roscommon Vol.1		1		1
Josie McDermott	Darby's Farewell	2	2		2
Catherine McEvoy	Flute Players of Roscommon Vol.1, Traditional Flute Playing in the Sligo-Roscommon Style	4			
Matt Molloy	Matt Molloy, Heathery Breeze, Shadows on Stone	5	2		
Conal O'Grada	Cnoc Bui	13	1	10	
Seamus Tansey	Field Recordings	4			
Michael Tubridy	The Eagle's Whistle	2	9		
John Wynne	Flute Players of Roscommon Vol.1		3		

Table 1: Dataset recordings showing player, album source and tune type.

gions from the feature matrix. The Adam optimiser is used to train the neural networks with an initial learning rate of 0.003. Training is stopped when the validation set accuracy does not increased between iterations. To ensure training commences correctly, the weights and biases are initialised to random non-zero values between ± 1 with zero mean and standard deviation equal to one. The performance measure used is cross entropy and the dropout probability d is set to 0.25 during training.

3. EVALUATION

As the performance of the proposed method depends heavily on the accuracy of the chosen onset detection method, the aim of our first evaluation is to determine the quality of existing timing data. We then perform an evaluation of our onset detection method by comparing it against the most successful method found in Ali-MacLachlan et al. (2016).

3.1 Dataset

The corpus for analysis consists of 79 solo flute recordings by nine prominent traditional flute players. Four common types of traditional Irish tune are represented: *reels*, *jigs*, *hornpipes* and *polkas*. Individual players are discussed in Köküer et al. (2014) and players, tune type and recording sources are detailed in Table 1.

The dataset contains annotations for onset timing information and labels for notes, cuts, strikes and breaths, and is comprised of approximately 18,000 individual events. First notes of long rolls, short rolls and cranns were also identified and labelled.

3.2 Onset detection evaluation

The ground truth annotation process was completed using multiple tools as the project evolved (Köküer et al., 2014; Ali-MacLachlan et al., 2015) resulting in inconsistencies being found in onset placement and labelling. We therefore improved the quality of these annotations by comparing ground truth onsets against true positive and false negative onsets obtained using `OnsetDetector` (Eyben et al., 2010). Events outside a $50ms$ window of acceptance were evaluated by an experienced flute player, allowing events to be checked for onset accuracy. Patterns containing impossible sequences of events were identified and eliminated by checking each event in context with previous and subsequent events.

To obtain the results for the `OnsetDetector` system on the updated dataset all tracks were processed with the output onset times compared against the annotated ground truth. We assess the accuracy relating to the `OnsetDetector` method before and after annotation correction and the number of spectrogram frames used as input.

We then evaluate the `OnsetDetector` system against the implemented CNN systems the dataset is divided by tracks into a 70% training set (55 tracks), 15% validation set (12 tracks) and 15% test set (12 tracks). The training set is used to train the five versions of the CNN (`CNN5`, `CNN11`, `CNN21`, `CNN41`, and `CNN101`) onset detector using the different input feature sizes, the validation set is used to prevent over-fitting and the test set is used as the unseen test data. The `OnsetDetector` results for the 12 test tracks are compared to the results from the 5 CNN versions. F-measure, precision and recall are used as the evaluation metrics with onsets being accepted as true positives if they fall within 25ms of the ground truth annotations.

4. RESULTS

4.1 Onset detection results

	P	R	F
`OnsetDetector` Before annotation improvement	83.06	75.10	78.75
`OnsetDetector` After annotation correction	85.86	78.46	81.85
`CNN5`	87.06	84.71	85.73
`CNN11`	88.07	84.73	86.25
`CNN21`	88.82	**88.26**	**88.46**
`CNN41`	**88.84**	86.63	87.58
`CNN101`	88.72	86.21	87.32

Table 2: Precision (P), Recall (R) and F-measure (F) for `OnsetDetector` (Eyben et al., 2010) before and after annotation improvement, `CNN5`, `CNN11`, `CNN21`, `CNN41`, and `CNN101`.

ISBN: 978-84-697-2303-6

			True Positives		
Label Code	Musical Pattern	Event Context	Onset Detector	`CNN21`	Total
`111`	note **note** note	*single notes*	1097	1124	1184
`211`	note **cut** note	*single cuts*	229	269	310
`121`	cut **note** note	*single cuts*	133	237	270
`112`	note **note** cut	*single cuts*	192	198	220
`114`	note **note** breath	*single notes*	96	99	106
`411`	note **breath** note	***single notes with breath***	21	53	88
`311`	note **strike** note	***single strike, end of roll***	55	42	76
`122`	cut **note** cut	*trill*	13	48	63
`141`	breath **note** note	*single notes*	55	56	61
`131`	strike **note** note	***single strike, end of roll***	16	33	57
`123`	cut **note** strike	*rolls*	14	33	36
`261`	note **cut** note	*start of long roll*	27	30	30
`153`	cut **note** strike	*start of short roll*	8	22	24
`511`	note **cut** note	*note before start of short roll*	18	21	23
`612`	note **note** cut	*note before start of long roll*	20	20	21
`142`	breath **note** cut	*breath before single cut*	19	20	20
`241`	breath **cut** note	*breath before single cut*	12	17	19
`412`	note **breath** cut	***breath before single cut***	3	11	19
`115`	note **note** cut	*two notes before start of short roll*	16	17	18
`271`	note **cut** note	*start of crann*	15	16	18
`116`	note **note** note	*two notes before start of long roll*	16	16	17
`113`	note **note** strike	*single strike*	14	13	15
`117`	note **note** note	*two notes before start of crann*	14	14	14
`712`	note **note** cut	*note before start of crann*	13	12	14
`132`	strike **note** cut	*cut after roll*	3	9	12

Table 3: Results comparing `OnsetDetector` and `CNN21` onset detectors for all event classes in the context of events happening prior and subsequent to the detected onset. Label codes of patterns with under 70% accuracy for `CNN21` shown in bold. Patterns with under 10 total onsets omitted.

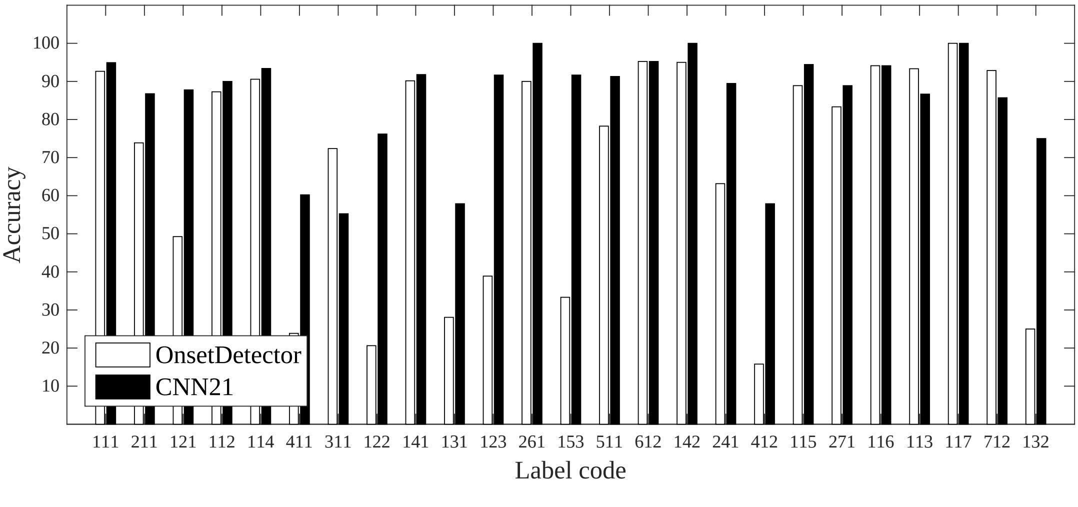


Figure 4: Accuracy of `OnsetDetector` and `CNN21` onset detectors for each event class above 10 onsets.

Table 2 presents the overall precision, recall and F-measure performance for the `OnsetDetector` and five CNN versions. The results indicate that all versions of the CNN achieve higher results than the `OnsetDetector`. The `CNN21`, which uses 10 spectrogram frames prior and subsequent to the middle frame achieves the highest recall and F-measure. The `CNN41` achieves a slightly higher precision than the `CNN21`, however achieves lower recall accuracy. The performance across the five CNN versions is fairly similar, illustrating that the moderate to higher values for the ψ parameter ($\psi = [21, 41, 101]$) are most appropriate for the task. The high performance of this approach is likely due to two factors. First, as CNNs are capable of processing large input feature sizes, they incorporate more context into the detection of a single frame. Second, as the CNNs are trained solely on traditional flute signals there is less variation in the represented classes, which has the potential of improving accuracy.

4.2 Note, cut and strike onset detection accuracy

Table 3 presents the onset detection results for each class of musical pattern with over 10 onsets in the test corpus of 12 tunes. The mean pattern precision across all classes was 79.22 for `CNN21` in comparison with 59.86 for `OnsetDetector`.

The classes consist of three event types where the central event is identified in bold. For example, label code 211 (*note* ***cut*** *note*) is a detected cut with a note before it and note after it, which exists within the event context of short and long roll or a single cut. The number of correctly detected onsets (true positives) is found as a percentage of the overall number of annotated onsets of that pattern. Label codes with an accuracy of less that 70% are shown in bold.

	Notes	Cuts	Strikes	Breaths
`OnsetDetector`	76.31	77.78	72.37	19.83
`CNN21`	89.57	91.29	55.26	59.06

Table 4: Accuracy of `OnsetDetector` and `CNN21` onset detectors for note, cut, strike and breath classes above 10 onsets.

As can be seen in Figure 4 and Table 3, low accuracies were found for strikes and notes following strikes. As a strike is played by momentarily tapping a finger over a tonehole, the pitch deviation is often much smaller than that of a cut and the event time is often shorter, making it more difficult to detect. Breaths are also difficult to detect in commercial recordings because it is usual to apply a generous amount of reverb effect at the mixing stage, resulting in a slow release masking a defined offset. Table 4 further illustrates inaccuracies in the detection of strikes and breaths by showing the accuracy for each single event class - note, cut, strike and breath. The note class also includes the notes at the start of ornaments such as long roll and crann and the cut class includes cuts at the start of short rolls.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented an onset detection method based a convolutional neural network (CNN) and is trained solely on Irish flute recordings. The results from the evaluation show that this method outperformed the existing state-of-the-art generalised trained `OnsetDetector`. We have also improved the annotations of a ITM dataset by employing a process of automatic onset detection followed by manual correction as required. To evaluate the effectiveness of this approach, the top performing CNN version (`CNN21`) method is compared to the `OnsetDetector` by (Eyben et al., 2010), most successful method found in Ali-MacLachlan et al. (2016).

In future research, we aim to develop note and ornament classification methods with additional features and attempt other neural network architectures in order to capture trends that appear in time-series data. We plan to release a corpus of solo flute recordings that will allow a deeper study into differences in playing style, and to extend this corpus to include other instruments. We also plan to investigate the generality of the proposed system to other instruments characterised by soft onsets such as the tin whistle and fiddle. The dataset used in this paper will also be released shortly, alongside Köküer et al. (2017).

6. REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., & others (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *CoRR*.

Ali-MacLachlan, I., Köküer, M., Athwal, C., & Jančovič, P. (2015). Towards the identification of Irish traditional flute players from commercial recordings. In *Proceedings of the 5th International Workshop on Folk Music Analysis*, Paris, France.

Ali-MacLachlan, I., Tomczak, M., Southall, C., & Hockman, J. (2016). Note, cut and strike detection for traditional Irish flute recordings. In *Proceedings of the 6th International Workshop on Folk Music Analysis*, Dublin, Ireland.

Beauguitte, P., Duggan, B., & Kelleher, J. (2016). A Corpus of Annotated Irish Traditional Dance Music Recordings: Design and Benchmark Evaluations.

Böck, S. & Widmer, G. (2013). Local Group Delay Based Vibrato and Tremolo Suppression for Onset Detection. In *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 589–594)., Curitiba, Brazil.

Boullier, D. (1998). *Exploring Irish Music and Dance*. Dublin, Ireland: O'Brien Press.

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.

Eyben, F., Böck, S., Schuller, B., & Graves, A. (2010). Universal Onset Detection with Bidirectional Long Short-Term Memory Neural Networks. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 589–594)., Utrecht, Netherlands.

Gainza, M., Coyle, E., & Lawlor, B. (2004). Single-note ornaments transcription for the irish tin whistle based on onset detection. *Proc Digital Audio Effects (DAFX), Naples*.

Gainza, M., Coyle, E., & Lawlor, B. (2005). Onset detection using comb filters. New Paltz, New York, USA.

Hast, D. E. & Scott (2004). *Music in Ireland: Experiencing Music, Expressing Culture*. Oxford, UK: Oxford University Press.

Ioffe, S. & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, *abs/1502.03167*.

Jančovič, P., Köküer, M., & Baptiste, W. (2015). Automatic transcription of ornamented Irish traditional music using Hidden Markov Models. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 756–762)., Malaga, Spain.

Keegan, N. (2010). The Parameters of Style in Irish Traditional Music. *Inbhear, Journal of Irish Music and Dance*, *1*(1), 63–96.

Kelleher, A., Fitzgerald, D., Gainza, M., Coyle, E., & Lawlor, B. (2005). Onset detection, music transcription and ornament detection for the traditional irish fiddle. In *Proceedings of the 118th AES Convention*, Barcelona, Spain.

Köküer, M., Ali-MacLachlan, I., Jančovič, P., & Athwal, C. (2014). Automated Detection of Single-Note Ornaments in Irish Traditional flute Playing. In *Proceedings of the 4th International Workshop on Folk Music Analysis*, Istanbul, Turkey.

Köküer, M., Ali-MacLachlan, Islah, Kearney, Daithi, & Jančovič, P. (2017). Curating and annotating a collection of traditional Irish recordings to facilitate stylistic analysis. *Special issue of the International Journal of Digital Libraries (IJDL) on Digital Libraries for Musicology, under review.*

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 *Proceedings of the 1st International Workshop on Digital Libraries for Musicology*, London.

Larsen, G. (2003). *The essential guide to Irish flute and tin whistle*. Pacific, Missouri, USA: Mel Bay Publications.

McCullough, L. E. (1977). Style in traditional Irish music. *Ethnomusicology*, *21*(1), 85–97.

Schluter, J. & Böck, S. (2014). Improved musical onset detection with convolutional neural networks. In *Acoustics, speech and signal processing (icassp), 2014 ieee international conference on*, (pp. 6979–6983). IEEE.

Southall, C., Stables, R., & Hockman, J. (2016). Automatic drum transcription using bi-directional recurrent neural networks. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 591–597)., New York City, United States.

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(1), 1929–1958.

TUNE CLASSIFICATION USING MULTILEVEL RECURSIVE LOCAL ALIGNMENT ALGORITHMS

Chris Walshaw
Department of Computing & Information Systems,
University of Greenwich, London SE10 9LS, UK
c.walshaw@gre.ac.uk

ABSTRACT

This paper investigates several enhancements to two well-established local alignment algorithms in the context of their use for melodic similarity. It uses the annotated dataset from the well-known Meertens Tune Collection to provide a ground truth and the research aim to answer the question, to what extent do these enhancements improve the quality of the algorithms? In the results, recursive application of the alignment algorithms, applied to a multi-level representation of the melodies, is shown to be very effective for improving the accuracy of the classification of the tunes into families. However, the ideas should be equally applicable to music search and melodic matching.

1. INTRODUCTION

1.1 Background

In the field of music information retrieval an important topic, impacting on music search, melodic matching / clustering and tune classification, is the calculation of melodic similarity. This paper investigates several enhancements to two well-established local alignment algorithms in the context of their use for melodic similarity.

It builds on results established by Janssen *et al.*, [1], [2] and van Kranenburg *et al.*, [3], which suggest that alignment-based similarity measures provide some of the best results for matching melodic segments, or indeed whole melodies, in a corpus of folk songs.

It uses an annotated dataset from the Meertens Tune Collection, [4], to provide a ground truth with which to evaluate the quality of the enhancements and for that reason deals with classification of melodies into tune families. However, the algorithms should be equally applicable to music search, e.g. [5], and melodic matching, e.g. [6], where some of the ideas were originally presented.

1.2 Organisation

The paper is organised as follows: section 2 presents the baseline algorithms and four enhancements. Section 3 discusses the evaluation and section 4 presents the results. Conclusions and further work are discussed in section 5.

2. ALGORITHMIC VARIANTS

This section discusses the alignment-based algorithms tested for the classification problem. Initially the two baseline algorithms, local alignment and longest common substring, are outlined (section 2.1). Subsequently a number of enhancements are discussed, including length normalisation (section 2.2) and the rhythmic representation of the melody (section 2.3), as well as globalising and multilevel enhancements (sections 2.4 & 2.5).

2.1 Baseline Similarity Measures

2.1.1 Local alignment (LA)

Local alignment is a well-known technique originating from molecular biology. Given two strings it finds the optimal alignment for two sub-sequences of the originals. The algorithm does not require the aligned sub-sequences to match exactly and makes allowances for gaps and substitutions. For example the strings `***abcde**` and `*acfe****` (where the asterisks represent non-matching entries) could potentially be aligned between `a` and `e` with a gap at the `b` and the substitution of `d` for `f`. Gaps (otherwise known as insertions and deletions) and substitutions are penalised with weights.

The algorithm is known as local alignment (LA) because, unlike the global alignment algorithms which preceded it, mismatching sub-strings from either side of the alignment are not penalised (i.e. in the example, the string of non-matching entries, indicated by asterisks, could be arbitrarily long without changing the alignment score).

To compute the optimal local alignment for two strings of length m & n, an (m+1) x (n+1) score matrix A is constructed with the top row and left hand column initialised to zero. The remainder of the matrix is then filled using

$$A(i,j) = \max \begin{cases} A(i-1,j-1) + s(X_i, Y_j) \\ A(i,j-1) + W_{\text{gap}} \\ A(i-1,j) + W_{\text{gap}} \\ 0 \end{cases}$$

$$\text{where } s(X_i, Y_j) = \begin{cases} W_{\text{match}} & \text{if } X_i = Y_j \\ W_{\text{substitution}} & \text{if } X_i \neq Y_j \end{cases}$$

and where W_{match}, $W_{\text{substitution}}$ and W_{gap} represent the weights for a matching or substituted entry or a gap in the aligned sequences. The implementation discussed here follows Janssen *et al.*, [1], [2], and uses $W_{\text{match}} = 1$, $W_{\text{substitution}} = -1$ and $W_{\text{gap}} = -0.5$.

This algorithm was introduced by Smith & Waterman, [7]. In fact their original scheme is a little more computationally involved but the scheme above is widely used and is the variant tested by Janssen *et al.*

To calculate the alignment score, and hence the qualitative similarity, the above scheme suffices. However to determine the aligned sub-sequences (needed for recursion, section 2.4) a trace-back procedure is required. The trace-back is implemented by recording a matrix of DIAG, UP or LEFT pointers for every entry of the score matrix indicating where the maximum value originated. If the maximum value is zero an END pointer is stored.

ISBN: 978-84-697-2303-6

The trace-back starts at the pointer matrix entry corresponding to the maximum score found and then tracks back through the pointers, terminating when it reaches an END. Diagonal moves indicate contiguous values in the two aligned sub-sequences whilst left or up moves indicate a gap in one of them.

2.1.2 Longest Common SubString (LCSS)

The longest common substring algorithm also finds matched sub-sequences from two strings but requires the sub-sequences to match exactly with no gaps or substitutions. It operates in a very similar fashion to local alignment filling in an (m+1) x (n+1) matrix of alignment values. However, because there is no need to allow for gaps, no trace-back is required: the position of the maximum score in the matrix indicates the end of the longest common substring and the value of this entry gives its length.

2.1.3 Sub-sequence alignment

In fact it is easy to see that, if the local alignment weights $W_{\text{substitution}}$ and W_{gap} are sufficiently large so that gaps and substitutions can never occur, then the LCSS algorithm is just a special case of local alignment.

From here on, therefore, both algorithms, LA and LCSS, will be referred to collectively as sub-sequence alignment, the main distinction between the two being that LCSS produces exact matching aligned substrings, is faster to compute and requires less memory (there is no need to use a full matrix and a memory efficient version exists which just repeatedly swaps a pair of arrays, one containing the row under calculation and one containing the previous row). Conversely, LA is more computationally complex and more memory intensive, but will generally match longer strings.

Both algorithms can be used for melodic similarity by representing each melody as a sequence of pitches or intervals: here intervals are used (see section 3.2). Then, if using $W_{\text{match}} = 1$ for LA, the similarity measure , S_{XY}, that either algorithm calculates between a pair of melodies, X and Y, represents the length (the number of notes) of the sub-sequences aligned. However, in the case of LA there may also be penalty weights for gaps or substitutions (for example, the matching of `abcde` with `acfe` has a score of $1 - ½ + 1 - 1 + 1 = 1½$).

2.2 Normalisation

The first simple algorithmic variant is just the way that the raw similarity measure is normalised. In their papers, [1], [2], Janssen *et al.* normalise the similarity S_{XY} by dividing by the length of the query. In the context of their use of short melodic phrases to query a database of melodies, and since the longest query is normally much shorter than the shortest melody, this means that the similarity measure is effectively normalised by min(length(X), length(Y)), In addition, since the maximum value possible of S_{XY} is also min(length(X), length(Y)), i.e. the length of the shortest of the two sequences being compared, then S_{XY} gives a value between 0.0 and 1.0 (with 1.0 being returned when an exact match of the query is found within the melody being queried).

For phrase-based classification studied in [1], [2], this makes perfect sense; there is no expectation that the query will match the entire tune. However, for the tune-based classification discussed here, that is no longer true and so using the minimum length may no longer be appropriate. For example, consider matching the sequence `abc` with two other sequences, `abc` and `abcdef`. In both cases the raw similarity is 3 and using minimum length (also 3) to normalise, the normalised similarity is 3/3 = 1.0. However, it does seem unreasonable that the similarity is the same in both cases (particularly since the match with the first sequence is exact, whereas the sequence `abcdef` could be arbitrarily long without changing the result).

Alternatives are to normalise with the maximum length or the average length. For example consider `abcxyz` matching with `abcdef` and `abcdefghi`. Using minimum length the normalised similarity is 0.5 for both matches (3/6) which doesn't seem unreasonable. However, using either maximum or average length, a sequence with identical raw similarity (in this case 3) to two other sequences will be normalised as closer to the one which is of a similar length, and arguably this is more appropriate.

Because it is not immediately clear which normalisation to use, the experimentation tests all three empirically.

2.3 Representation – bar indicators

A second algorithmic variant tries to take account of the position of notes within the bar. This is particularly relevant for folk music (although perhaps less so for jazz) since the position often determines which are the stressed (more important) notes.

One way to achieve this is to adapt the similarity measure to add weight to stressed notes, e.g. [8]. However, that relies on what is arguably a subjective assessment of which are the stressed notes. Instead, as discussed in [5], it is possible to use bar indicators or even bar numbers in the sequences of intervals to be compared. For example a major scale can be indicated by the intervals `2212221`. Including bar indicators, and assuming 4 notes to the bar this could be represented as `|221|2221|` where the `|` symbols represent bar lines (note that an interval between the last note in a bar and the first note in the next bar, could be shown before or after the bar indicator; in the experiments here it is always included afterwards). 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. |$_1$221|$_2$2221|$_3$, where each |$_i$ represents a numbered bar, then matched common substrings also need to respect the position in the tune. (If matching of subsections of the tune is important then the numbering could be restarted at natural breaks such as double bar lines and repeat marks; however, that is not 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 indicators is more of a representational variant than an algorithmic one and increases the computational complexity of the matching slightly (as the strings to be compared by the similarity measure are longer). However, even though some melodies in the dataset under investigation are in free meter and have no bar lines, it has a significant effect on the results and is an important enhancement.

2.4 Recursive sub-sequence alignment

2.4.1 Recursive alignment (= global alignment)

A problem with using LCSS, and to a lesser extent LA, is that they are local. For example, using LCSS, `ab**ba` has exactly the same raw alignment score (of 2) when matched with `**ab` and with `ab**ba`, even though the latter seems a far better match. This is because the second match (`ba`) is not accounted for.

This was less of an issue in the predecessor to this paper, [9], where LCSS was used as part of a multilevel melodic search algorithm, since search algorithms are typically trying to find the best matches of a short phrase in a dataset of complete melodies. However, for classification it is crucial to distinguish between tunes which match well across their entire length and those which perhaps only match for a short segment.

Interestingly Smith & Waterman touch on this in their original paper where they say "the pair of segments with the next best similarity is found by applying the trace-back procedure to the second largest element of [the matrix] not associated with the first trace-back", [7]

Unfortunately, working from the existing matrix may lead to overlapping local alignments and instead sub-sequence alignment may be applied recursively as follows: when applied to two strings, S1 and S2, sub-sequence alignment splits both into three substrings S1 = L1 + A1 + R1 and S2 = L2 + A2 + R2, where A1 and A2 are the aligned substrings (exact matches for LCSS or potentially with gaps and substitutions for LA), L1 and L2 are the left hand side unmatched substrings and R1 and R2 are the right hand side unmatched substrings (where any of the these unmatched substrings may be of length 0). Thus, having found A1 & A2 and split S1 & S2, sub-sequence alignment can then be applied to compare L1 with L2 and R1 with R2.

This procedure continues recursively, terminating when no alignment is found, or one or both lengths of the substrings being aligned are 0. For example, if the start of S1 is aligned with the end of S2 no further recursion is possible as the lengths of L1 and R2 are 0.

This recursion effectively turns the local alignment algorithms LCSS or LA into a globalised similarity measure, giving an alignment score along the length of both strings being compared. Henceforth these recursive algorithms will be referred to as RLCSS and RLA.

2.4.2 Biased recursive local alignment

An issue that became apparent when using recursive alignment, is that just adding all the scores together makes no distinction between one long aligned sequence and several shorter ones. For example (using RLCSS) `abcd****` has the same alignment score (of 4) when compared with `abcd****` and with `**a**b**c**d**`, even though the former seems a good match and the matching with the latter is essentially noise.

To address this, the similarity measure can be biased towards longer aligned sub-sequences by taking the 2-norm (square root of the sum of squares) of the alignment scores found by the recursive local alignment. In the above example this means that the biased recursive local alignment score is $\sqrt{4^2} = 4$ when matching `abcd****` with `abcd****`, whereas when matching with `**a**b**c**d**` it is $\sqrt{1^2 + 1^2 + 1^2 + 1^2} = 2$.

2.5 Multilevel Similarity

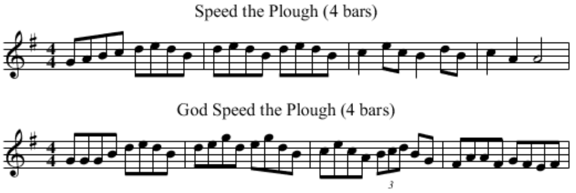


Figure 1. Two tune variants for Speed the Plough.

Multilevel similarity was first introduced in [5] and subsequently developed further in [6]. The idea is motivated in Fig. 1 which shows two versions of the first 4 bars of Speed the Plough, a tune well-known across the British Isles. Clearly these tunes are related but with distinct differences, particularly in the second and fourth bars.

It is typical in tunes like this 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. [8]). However, an alternative approach is to build a multilevel (hierarchical) representation of the tunes.

Figs. 3 & 4 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. (Interestingly this accords with an idea used by Breathnach, [10], the renowned collector of Irish traditional music, who developed a system for indexing melodies based on the accented notes of each tune – effectively level 1 in the multilevel hierarchy).

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.

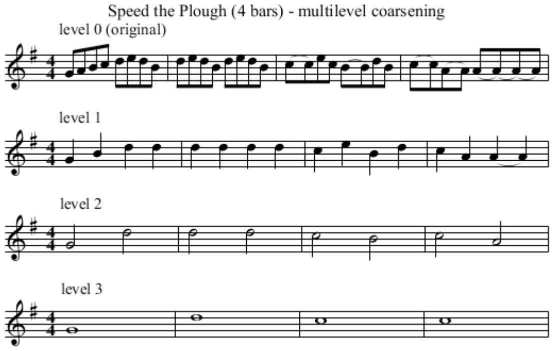


Figure 2. Multilevel coarsening of Speed the Plough

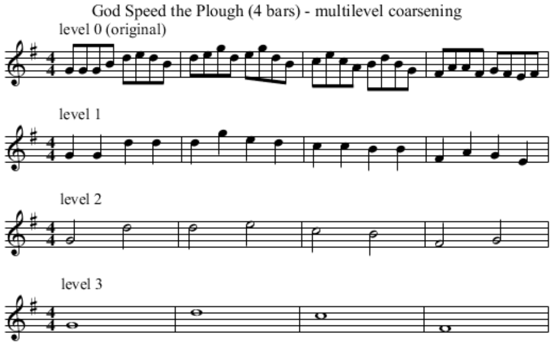


Figure 3. Multilevel coarsening of God Speed the Plough

As can be seen, as the coarsening progresses the two versions become increasingly similar and thus provide a good scope for melodic comparisons by ignoring the finer details of the tunes.

The implementation of this scheme is discussed more fully in [6] but is mostly straightforward. Each tune is initially normalised & quantised and then recursively coarsened down to a skeleton representation with just one note per bar. Melodic similarity calculations can then take place at every level and the (possibly weighted) sum of the similarities at each level used to provide a multilevel similarity measure (see section 4.5).

The coarsening works by recursively removing "weaker" notes from each tune to give increasingly sparse representations of the melody. In the current implementation the coarsening strategy considers that the weaker notes are the off-beats or every other note and it is these which are removed (see Figs. 3 & 4). However, it should be stressed that the multilevel framework is not tied to a particular coarsening strategy and in principle any algorithm that can reduce the detail in the melody (preferably recursively) could be used. For example, it should even be possible to use something as complex as a Schenkerian reduction, [11]; conversely multilevel algorithms in other fields often use randomised coarsening, [12].

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. The same 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, are discussed in [9].

Coarsening progresses until there is one note remaining in each bar; it would be possible to take it further, coarsening down to one single note for a tune, but experimentation suggests that the bar is a good place to stop. In fact some tunes in the dataset under investigation are in free meter, with no bar lines, and hence are coarsened down to a single note. An artificial limit of 4 levels (typical for many time signatures) was tested, but made very little difference to the results, particularly since any excess levels are ignored when comparing melodies with differing numbers of levels.

Once the multilevel representation is constructed, a variety of methods (including the alignment algorithms discussed here) can be used to compare each level. Again, this is a strength of the multilevel paradigm which is not reliant on a particular local search strategy, [12].

In the experiments below, multilevel variants are referred to as ML-*, where the * indicates the sub-sequence alignment algorithm (e.g. RLA). Conversely, if the multilevel representations are not used the similarity framework is referred to as SL-* (i.e. single level).

3. EVALUATION

3.1 The dataset

The dataset used is the Annotated Corpus of the Meertens Tune Collection, version 2.0.1, [4]. This contains 360 melodies each identified as belonging to one of 26 tune families. It also includes further annotations, splitting each melody into phrases, with three annotators manually assigning labels to each phrase. These have been used by Janssen *et al.*, [1], [2], for testing search queries based on phrases, rather than whole melodies. However, since the investigation under discussion deals with globalised a2lignment algorithms it was decided to ignore the breakdown of the melodies into their constituent phrases.

3.2 Representation - transposition & time dilation

In [2], Janssen *et al.* present a comparison of several algorithms and indicate that what can make a big difference to the results is the representation of the music. In particular they find that using pitch adjustment in order to resolve transposition differences (i.e. similar melodies transcribed in different keys) can significantly improve the performance of the algorithms. However, the premise for their research is that the tune families are known in advance and so the pitch adjustment scheme uses this information and aims to transpose all the melodies in a given family into the same key. This does not apply for the work presented here where the aim is to classify each query melody into one of the 26 tune families, under the assumption that this is not known beforehand.

Perhaps a more appropriate scheme would be the pairwise pitch adjustment used by van Kranenburg *et al.*, [3]. However, in the experimentation below, the algorithms are made transposition invariant by representing each melody as a sequence of pitch intervals. In contrast with Janssen *et al.*, [2], this was not found to deteriorate the algorithmic performance and it may be the case that their pitch adjustment scheme provides better results than intervals because it uses tune family information that would not normally be available for an arbitrary dataset.

Another interesting representational idea is the duration adjustment scheme, again from Janssen *et al.*, [2], which seeks to adjust the note durations in a similar manner to pitch adjustment so that melodies transcribed in different meters (e.g. 3/4 and 6/8) are more closely comparable. However, that has not yet been tested with the algorithms described here.

In terms of the variants discussed by Janssen *et al.*, the musical representation used in all experiments presented here is duration weighted pitch intervals, i.e. a representation which repeats each pitch according to the length of the note and with a sequence of integers expressing the difference in semitones between successive pitches.

3.3 Evaluation – Receiver Operating Curves (ROC)

The evaluation of each algorithmic variant is straightforward. Each of the 360 melodies is used as a query and compared against the other 359 melodies with the algorithm assigning a similarity score between the query and each melody. This results in 360 arrays, each containing 359 similarity scores.

Each array can then be used to generate a Receiver Operating Characteristic (ROC) curve which plots the true positive rate (TPR) against the false positive rate (FPR) in the ground truth as the results array is traversed. ROC curves are an elegant, generic tool often used to evaluate classification experiments across a wide range of disciplines, [13]. In fact they do not even require the similarity scores as input, they just need the results sorted in order of decreasing similarity and the ground truth (in this context whether a melody belongs to the same tune family as the query or not) to determine positive or negative outcomes.

Typically ROC curves are compared by measuring the Area Under the Curve (AUC). Since any ROC is confined to the unit square, the corresponding AUC is a value between 0.0 and 1.0 with higher values indicating a better classification algorithm. An AUC value of 1.0 indicates that the algorithm has done a perfect classification with all the true positives sorted by the similarity scores to one end of the array (and hence all the true negatives sorted to the other end). Conversely an AUC of 0.5 indicates that the algorithm has essentially done no better than a random classification.

Since each algorithmic variant results in 360 ROC curves, a method for combining them together is required. Janssen *et al.*, [1], [2], aggregate all of the similarity results into one ROC curve and then measure the area underneath. However although this is a recognised technique, e.g. [13], it is not area-preserving in the sense that the average area under the individual curves is not necessarily the same as the area under the aggregated curve.

To see this, suppose an algorithm produces similarity scores of [1.00, 0.49, 0.00] for a particular search query and dataset of 3 melodies when the corresponding ground truth is [true, true, false] (in other words the melody with the similarity score of 0.00 is not a member of the same family as the search query, whereas the other two are). Since the similarity measure has done a perfect job of ordering the dataset using the similarity scores (perfect in the sense that all the true matches are at left hand end of the array and all the false matches at the other end), the ROC curve representing this would actually run up the x-axis and then along the line y = 1, giving the maximum possible AUC of 1.0.

Now suppose that a second search query produces similarity scores of [1.00, 1.00, 0.51] with the same corresponding ground truth of [true, true, false]. Once again the ordering is perfect and the area under the curve is 1.0. So the average AUC across the 2 queries is 1.0.

However, if the scores and ground truths are aggregated to form a single curve the results are no longer perfect as 0.49 is smaller than 0.51 and so the ordered ground truth array is [true, true, true, false, true, false]. The AUC for the corresponding ROC is 0.875.

Conversely, consider 2 search queries used on a dataset of 4 melodies and producing the results [1.00, 1.00, 0.52, 0.51] and [1.00, 1.00, 0.49, 0.48], both with corresponding ground truth of [true, true, false, true]. In this case the classifier has not done a perfect job and the AUC for each ROC is 0.667. When the results are aggregated the ordered ground truth array is [true, true, true, true, false, true, false, true] and the corresponding aggregated ROC has an AUC of 0.75.

Thus it is possible that the area under the aggregated curve can be significantly different (either lower or higher) from the average area under the individual curves.

Of course, as more results (more search queries, a larger dataset) are included, it is likely that the differences between the average AUC and the AUC for the aggregated ROC will diminish. Nonetheless, the aggregated ROC may not be telling the whole story.

In this paper the results for each algorithm are aggregated simply by taking an average of all the AUC values for that algorithm. Then, in order to draw the ROCs in Fig. 1, it is possible to use Fawcett's vertical averaging algorithm in [13] (Algorithm 3). Although Fawcett describes vertical averaging by sampling the ROC space at regular intervals, this is easily adapted to the non-parametric scheme described by Chen & Samuelson, [14], where it is sampled at every possible FPR value. With this adaptation in place, Chen & Samuelson have proven that the averaged ROC is area preserving, i.e. the AUC for the averaged ROC is the same as the average AUC across the individual ROCs (up to rounding differences).

3.4 Classification success rate (CSR)

Finally, to evaluate the quality of the tune classification into families, the nearest neighbour scheme described by van Kranenburg *et al.*, [3], is applied. Specifically the melody in use567 as a query is assigned to the tune family of the nearest neighbour in terms of similarity. This assumes that the tune families of the 359 other melodies are known and that of the query is the unknown. However, for an arbitrary unannotated dataset, with no known tune families, it should be possible to use proximity graphs, similar to those described in [6] and with suitably chosen thresholds, to suggest tune family membership.

In the event that a number of melodies are nearest neighbours (i.e. have the same similarity with the query) then here ties are broken by considering the set of all such melodies and picking the tune family with the largest similarity contribution across the set.

Finally since, unlike van Kranenburg *et al.*, [3], this experimentation is only applied to the small annotated dataset of 360 tunes, the classification success rate (CSR) can be calculated as a simple percentage |S|/360 where S is the set of tune family labels successfully identified.

4. EXPERIMENTATION

This section discusses the results: throughout the algorithms are applied cumulatively with the best perorming approach from each section used in the following section.

4.1 Baseline results

Table 1 shows the results for the baseline algorithms, single level LCSS & LA, showing the average AUC across all of the queries (which as mentioned above is the same as the AUC for the averaged ROC) and the classification success rate (CSR).

Algorithm	Variant	Avg AUC	CSR
SL-LCSS	baseline	**0.787**	0.697
SL-LA	baseline	**0.787**	0.814

Table 1. Results from the baseline algorithms, LCSS & LA (see section 2.1).

In this table (as all others) the best AUC figures for LCSS & LA variants are shown in boldface to highlight the key performance indicator.

What is perhaps surprising is that the LCSS algorithm appears to perform as well as the LA algorithm although it does a worse job of classification (69.7% correct as compared with 81.4%). However, the average AUC is 0.787 for both algorithms indicates a generally high quality similarity measure and, although the figures cannot be directly compared (for the reasons given in sections 3.1, 3.2 & 3.3), is broadly comparable to the 0.790 figure for LA in [1].

It should not be a surprise that there is such a wide difference in the Classification Success Rate (CSR). In fact CSR is not such a good performance indicator as AUC, since essentially it only applies to the nearest neighbour (highest similarity) for each query, whereas the AUC measures the performance of the similarity measure across the entire dataset. Thus, as well as indicating the similarity with all other melodies in the tune family, the AUC is a better indicator of how the algorithm might perform for other melodic similarity tasks, such as search and matching.

4.2 Length normalisation

Table 2 shows the effect of applying different length normalisations to the similarity measure – i.e. when comparing two tunes of different lengths, dividing the raw similarity score by the minimum, the average and the maximum length of the two tunes (of course, if the tunes are the same length then these three values are the same).

As can be seen, this can make a small improvement to the results, with minimum length giving the worst results and average length the best. Surprisingly here, LCSS even outperforms LA.

Algorithm	Length	Avg AUC	CSR
SL-LCSS	Min	0.787	0.697
SL-LA	Min	0.787	0.814
SL-LCSS	Avg	**0.818**	0.853
SL-LA	Avg	**0.810**	0.872
SL-LCSS	max	**0.818**	0.872
SL-LA	max	0.802	0.883

Table 2. Results showing the effects of different length normalisation (see section 2.2).

From here on all results use average length as the chosen normalisation, unless otherwise indicated

4.3 Bar indicators

Table 3 shows the effect of including bar indicators in the representation. As can be seen, bar markers and even bar numbers can improve some results significantly, with bar numbers being somewhat less effective. This is perhaps to be expected; bar numbers tie the bars down to a particular part of the melody whereas in fact there are known instances where the ordering of the phrases may change.

Algorithm	Bar indicators	Avg AUC	CSR
SL-LCSS	none	0.818	0.853
SL-LA	none	0.810	0.872
SL-LCSS	markers	**0.827**	0.853
SL-LA	markers	**0.849**	0.911
SL-LCSS	numbers	0.814	0.853
SL-LA	numbers	0.846	0.906

Table 3. Results showing the effects of using bar indicators (see section 2.3).

Furthermore, with bar markers and average length normalisation LA is now seen to give better results than LCSS. Again this is to be expected since it is a more sophisticated (though computationally costly) algorithm.

4.4 Recursive sub-sequence alignment

Table 4 shows the effect of including recursive variants of the sub-sequence alignment algorithms, i.e. Recursive Local Alignment (RLA) and Recursive Longest Common SubString (RLCSS). As can be seen, the crucial feature is the use of biased similarity (biased to favour longer matches, rather than a series of short matches) which uses the 2-norm of the recursive similarity scores (section 2.4.2); just adding the recursive similarity scores together (1-norm) actually makes the recursive results worse than the non-recursive versions.

Algorithm	Recursive score	Avg AUC	CSR
SL-LCSS	none	0.827	0.853
SL-LA	none	0.849	0.911
SL-RLCSS	1-norm	0.813	0.828
SL-RLA	1-norm	0.842	0.878
SL-RLCSS	2-norm	**0.845**	0.889
SL-RLA	2-norm	**0.854**	0.914

Table 4. Results showing the effects of using recursive sub-sequence alignment (see section 2.4).

4.5 Multilevel similarity

Table 5 presents perhaps the biggest performance enhancement which comes from the multilevel similarity measure, adding all the similarity scores from all coarsened versions of the melody. This significantly improves on the single level versions, SL-RLCSS and SL-RLA.

Algorithm	Framework	Avg AUC	CSR
SL-RLCSS	single level	0.845	0.889
SL-RLA	single level	0.854	0.914
ML-RLCSS	multilevel	**0.870**	0.900
ML-RLA	multilevel	**0.887**	0.922

Table 5. Results showing the effects of using multilevel similarity (see section 2.5).

Note that other experiments were performed to vary the weight of the similarity contributions from each level (e.g. as suggested in [5], giving greater weight to the finer, more accurate representations of the melody). However, none of the variants gave consistently better results.

4.6 Parameter cross-checking

Finally Tables 6 & 7 provide some cross-checks to further validate the results above. Table 6 shows the results for the multilevel schemes, using bar markers but with different length normalisations (minimum, average & maximum). As in section 4.2, the average length is seen to give the best normalisation.

Algorithm	Length	Avg AUC	CSR
ML-RLCSS	Min	0.840	0.811
ML-RLA	Min	0.865	0.872
ML-RLCSS	Avg	**0.870**	0.900
ML-RLA	Avg	**0.887**	0.922
ML-RLCSS	max	0.866	0.900
ML-RLA	max	0.878	0.917

Table 6. Results showing the effects of different length normalisation for the multilevel algorithms.

Meanwhile Table 7 shows the results using average length normalisation, but comparing bar indicators (no indicators, bar markers & bar numbers). As in section 4.3, bar markers are seen to give the best results.

Algorithm	Indicators	Avg AUC	CSR
ML-RLCSS	none	0.880	0.908
ML-RLA	none	0.883	0.922
ML-RLCSS	bar markers	**0.870**	0.900
ML-RLA	bar markers	**0.887**	0.922
ML-RLCSS	bar numbers	0.849	0.883
ML-RLA	bar numbers	0.868	0.925

Table 7. Results showing the effects of using bar indicators for the multilevel algorithms.

4.7 Discussion

Fig. 1 shows the ROC curves for 4 of the algorithmic variants, the two baseline algorithms (SL-LCSS & SL-LA) and the two final algorithms with all four enhancements (ML-RLCSS & ML-RLA using average length normalisation and bar markers). As can be seen the baseline algorithms have very similar curves with SL-LCSS marginally worse than SL-LA for smaller values of TPR / FPR (i.e. with high similarities) and marginally better at the other end of the range. Of the final algorithms, ML-RLA is better than ML-RLCSS although their performance is almost indistinguishable for larger values of TPR / FPR.

Note also that although LA versions of the algorithms generally outperform LCSS, the LCSS results are of interest because they achieve nearly the same quality and are much faster (e.g. in the tests presented here LCSS variants are about 1.4 to 1.6 times faster than the LA counterparts).

As mentioned before, the results here are not directly comparable with those of Janssen *et al.*, [1], [2] (for the reasons given in sections 3.1, 3.2 & 3.3). Nonetheless they are of the same order and for example the best AUC value presented here (0.887), even without knowledge of the tune families, is broadly comparable to the best value in [2] (0.893, achieved using a hand-adjusted representation). It is even possible that by combining some of the techniques (e.g. the duration adjustment scheme, [2]) the results could be improved still further.

The results are more directly comparable with those of van Kranenburg *et al.*, [3]. Unfortunately the best CSR of 0.925 presented here does not match the CSR of 0.99 achieved there and again this argues for further integration of techniques.

5. CONCLUSIONS

This paper has investigated several enhancements to two well-established sub-sequence alignment algorithms, in the context of their use for melodic similarity and in particular classification of queries into tune families. It uses the annotated dataset from the well-known Meertens Tune Collection to provide the ground truth with which to evaluate the quality of the algorithms.

In particular, recursive application of the alignment algorithms applied to a multilevel representation of the melodies is shown to be very effective for improving the accuracy of the classification.

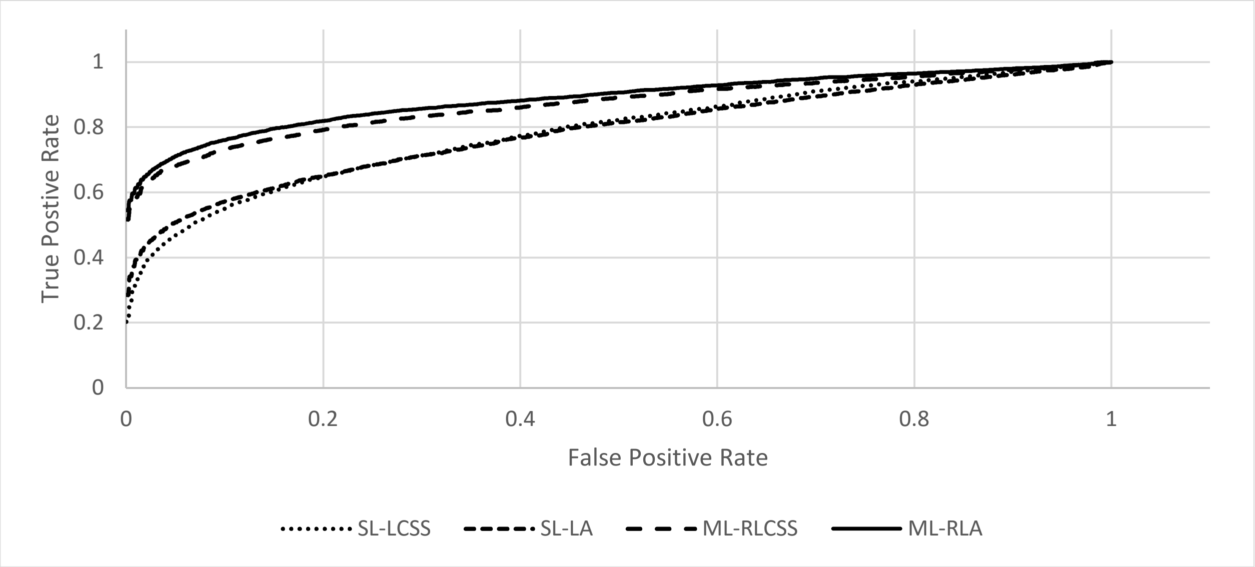


Figure 4. The averaged ROC curves for four algorithmic variants: the baseline algorithms and the best variants.

The other enhancements include length normalisation of the similarity measure (which can be tailored according to the problem – for example minimum length might be more appropriate where all the queries are expected to be short phrases but the dataset contains complete melodies). In addition, the use of bar indicators can improve the results still further.

In broad terms, the impact of these enhancements (along with other representational variants, e.g. [2]) suggest that sub-sequence alignment (both the local alignment version, LA, and the special case, LCSS) are flexible and robust in terms of how the music is represented and how the algorithms are applied.

Finally it should be stressed that these enhancements do not appear to be mutually dependent. In other words, it should be possible for other authors to adopt some or all of the algorithmic enhancements discussed here to improve melodic similarity algorithm(s), and the ideas should be equally applicable to music search and melodic matching.

6. REFERENCES

[1] B. Janssen, P. van Kranenburg, and A. Volk, “A Comparison of Symbolic Similarity Measures for Finding Occurrences of Melodic Segments,” in *Proc. ISMIR*, 2015, pp. 659–665.

[2] B. Janssen, P. van Kranenburg, and A. Volk, “Finding occurrences of melodic segments in folk songs employing symbolic similarity measures,” *J. New Music Res.*, p. (to appear), 2017.

[3] P. van Kranenburg, A. Volk, and F. Wiering, “A Comparison between Global and Local Features for Computational Classification of Folk Song Melodies,” *J. New Music Res.*, vol. 42, no. 1, pp. 1–18, 2013.

[4] P. van Kranenburg, B. Janssen, and A. Volk, “The Meertens Tune Collections : The Annotated Corpus (MTC-ANN) Versions 1.1 and 2.0.1,” 2016.

[5] C. Walshaw, “Multilevel Melodic Matching,” in *5th Intl Workshop on Folk Music Analysis*, 2015, pp. 130–137.

[6] C. Walshaw, “Constructing Proximity Graphs To Explore Similarities in Large-Scale Melodic Datasets,” in *6th Intl Workshop on Folk Music Analysis*, 2016.

[7] T. F. Smith and M. S. Waterman, “Identification of common molecular subsequences,” *Mol. Biol.*, vol. 147, pp. 195–197, 1981.

[8] R. Typke, “Music Retrieval based on Melodic Similarity,” Utrecht University, Netherlands, 2007.

[9] C. Walshaw, “TuneGraph: an online visual tool for exploring melodic similarity,” in *Proc. Digital Research in the Humanities and Arts*, 2015, pp. 55–64.

[10] B. Breathnach, “Between the Jigs and the Reels,” *Ceol*, vol. V, no. 2, pp. 43–38, 1982.

[11] A. Marsden, “Schenkerian Analysis by Computer: A Proof of Concept,” *J. New Music Res.*, vol. 39, no. 3, pp. 269–289, 2010.

[12] C. Walshaw, “Multilevel Refinement for Combinatorial Optimisation: Boosting Metaheuristic Performance,” in *Hybrid Metaheuristics - An emergent approach for optimization*, C. Blum, Ed. Springer, Berlin, 2008, pp. 261–289.

[13] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognit. Lett.*, vol. 27, pp. 861–874, 2006.

[14] W. Chen and F. W. Samuelson, “The average receiver operating characteristic curve in multireader multicase imaging studies,” *Br. J. Radiol.*, vol. 87, no. 1040, pp. 1–8, 2014.

IMAGE-BASED SINGER IDENTIFICATION IN FLAMENCO VIDEOS

Nadine Kroher
University of Seville
nkroher@us.es

Aggelos Pikrakis
University of Piraeus
pikrakis@unipi.gr

José-Miguel Díaz-Báñez
University of Seville
dbanez@us.es

ABSTRACT

Automatic singer identification is an essential tool for the organization of large poorly annotated music collections. For the particular case of flamenco music, we identify various common scenarios where the singer label is either incorrect or missing. We propose an image-based singer identification method for flamenco videos using state of the art face recognition technologies. First, we detect faces in video-frames using the HOG method. We then determine if a face belongs to the singer by analyzing the mouth opening. From all faces which are associated with the singer, we generate an embedding using a pre-trained deep convolutional network and evaluate against a database containing labeled singer images. In an experimental setup we obtain a classification accuracy of 90% which is a promising result compared to an audio-based baseline method and considering the diversity of quality and recording scenarios contained in the database.

1. INTRODUCTION

The technologically challenging task of automatic singer identification is of crucial importance for the automatic indexing of large music databases. For the particular case of flamenco music there are several frequently occurring scenarios where the singer in a performance is unknown: One example can be found in performance videos starring renown dancers, where singers are usually considered in an accompanying role and in many cases only the name of the dancer is annotated. However, many respected singers spent the early years of their career accompanying dancers and discovering such videos could be beneficial for studying the evolution of a singer over time. Furthermore, names are often not unique identifiers in the flamenco world. Singers may be referred to by both their stage name as well as their actual name and related singers, i.e. father and son, may be referred to by the same name. In addition, more and more flamenco videos are submitted to popular multi-purpose video sharing platforms. However, such platforms do often not require to annotate the artist performing in the video. As a result, many flamenco videos are labeled by genre or style only.

Related work on automatic singer identification has so far been limited to the analysis of audio recordings. Most approaches have used machine learning models trained on low-level timbre descriptors (Cai et al., 2011; Tsai & Lee, 2012; Lagrange et al., 2012; Shen et al., 2009; Zhang, 2003), fundamental frequency trajectories (Fujihara et al., 2010) or vibrato-related descriptors (Nwe & Li, 2008). Methods for non-Western music traditions addressing genre-specific properties and challenges have been developed for carnatic music (Sridhar & Geetha, 2008), rembetiko (Holzapfel & Stylianou, 2007) and flamenco (Kroher & Gómez, 2014). However, spectral distortions in low audio quality audio recordings and the presence of dominant accompaniment instruments have shown to limit the performance of audio-based approaches.

Motivated by the growing amounts of digitally available audio-visual performance recordings and the challenges described above, we present an image-based approach to singer identification in flamenco videos using state of the art face recognition technologies. The term face recognition refers to the task of automatically identifying a person based on a facial image. Given their non-intrusive nature and low-cost hardware requirements, face recognition methods are an essential tool for biometric-based person identification (Moon, 2004) and surveillance (Burton et al., 1999), and have furthermore found application in multimedia indexing and video thumbnailing (Lee, 2005). For a compete review we refer the reader to Jafri & Arabnia (2009).

The task of singer identification in music performance videos encompasses two major challenges: First, we need to detect the face of the singer despite the presence of various musicians on stage. Then, we need to determine the singer's identity among a number of candidates in an annotated image database. In Section 2 the method is described in detail, an overview of the dataset used in this study is given in Section 3 and experimental results are provided in Section 4. The paper is concluded in Section 5

2. METHOD

An overview of the proposed method is depicted in Figure 1. First, we train a machine learning model on a set of annotated frontal images of singers. In each image, we detect the face bounding box using a state of the art face detection algorithm and then align all faces to a canonical pose. Subsequently, we extract a vector of discriminatory features, called *face embedding*, from all images. In order to identify the singer of an unlabeled video file, we first detect all face bounding boxes in each frame. We then decide if a detected face corresponds to the singer, by extracting face landmarks and estimating the amount of mouth opening. If the mouth is estimated to be open, we assume that the face inside the bounding box belongs to the singer and proceed as in the training stage: We align the face image and extract its embedding. We compute pair-wise similarities to all instances in the training set and then classify based on the labels of the most similar images. Finally, in order to assign a label to the video file, we perform a weighted voting scheme over all frame-wise estimates and their confidence values. Below, all processing stages are

ISBN: 978-84-697-2303-6

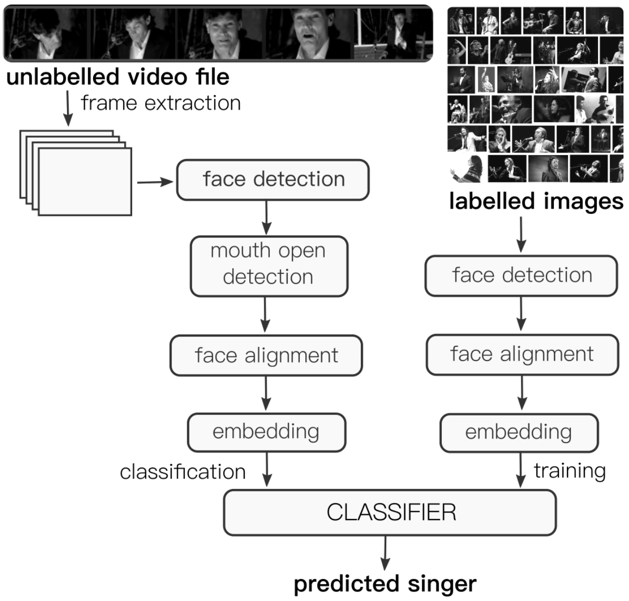


Figure 1: Overview of the processing pipeline.

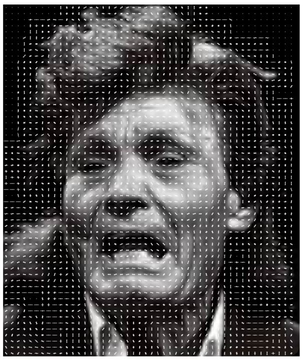


Figure 2: Image with overlay of its HOG representation.

described in detail.

2.1 Face detection

We apply the histogram of oriented gradients (HOG) method introduced by Dalal & Triggs (2005) in order to detect faces in an image. The HOG representation of an image is generated by computing the brightness gradient for each pixel and then determining the dominant gradient in cells of 16x16 pixels. An example is shown in Figure 2. Using a sliding window with multiple scales, the local HOG representation can then be evaluated against a pre-trained model. Here, we used the implementation available in the *dlib* library (King, 2009) together with a linear classifier pre-trained on the *labeled faces in the wild* (Learned-Miller et al., 2016) dataset.

2.2 Landmark estimation and alignment

In order to extract pose-invariant features, it is necessary to align all images to a reference pose. To this extent, we first crop the image at the estimated bounding box and apply the method proposed by Kazemi & Sullivan (2014) to detect 68 facial landmarks. Subsequently, a number of affine transforms are performed to shift the landmarks corresponding to outer eyes and nose to a reference position. The method is implemented in the *dlib* library (King, 2009). An example of face detection, landmark estimation and alignment is shown in Figure 3.

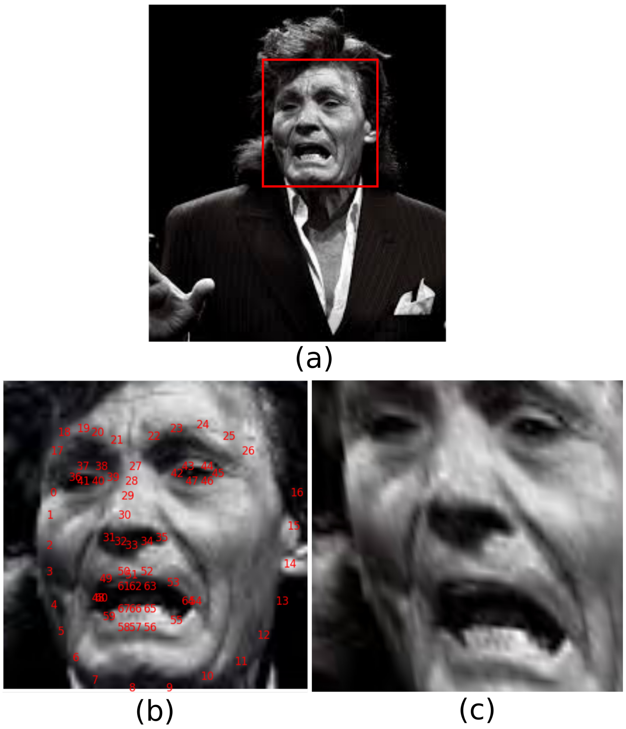


Figure 3: (a) original image with face bounding box; (b) cropped image and face landmarks; (c) cropped and aligned image.

2.3 Mouth open detection

In flamenco videos we often encounter, apart from the singer, various musicians on stage, including guitarists, dancers and percussionists. Consequently, it is necessary to decide if a detected face belongs to the singer. Here, we assume that detected faces with a wide mouth opening are most likely frontal shots of the singer. Therefore, we compute the relative distance d between the estimated facial landmarks corresponding to the center of upper and lower lip, l_{up} and l_{low} respectively, with respect to the height of the face bounding box h_{face}:

$$d = \frac{l_{up} - l_{low}}{h_{face}} \tag{1}$$

An example for $d = 0.2$ is shown in Figure 4. We experimentally examined the value of d during various singing passages and determined $d > 0.15$ as a hard threshold for detecting a mouth to be open.

2.4 Embedding

The training and evaluation of a machine learning model for face recognition requires the extraction of representative features with high discriminatory power. Here, we ex-

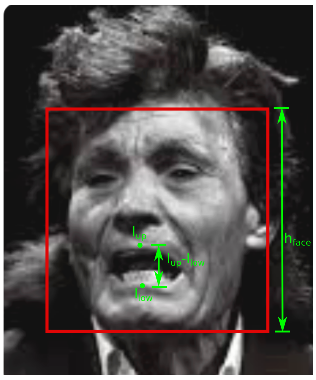


Figure 4: Example of an estimated mouth opening of $d = 0.2$

tract the so-called face embedding as proposed by Schroff et al. (2015). This mapping of an input image to 128 features was learned using a deep convolutional neural network on the faceSCRUB (Ng & Winkler, 2014) and CASIA webFace (Yi et al., 2014) datasets. Here, we use the implementation which is available together with the pretrained model in the *openFace* library (Ambos et al., 2016).

2.5 Classification

Given a raw video file, we extract image frames in intervals of one second and generate the embedding of faces for which the mouth was estimated to be open. We evaluate each of these feature sets against the embeddings of a labeled database in a weighted k-nearest neighbor (k-NN) classification scheme (Fix & Hodges Jr, 1951). We chose the rather simple k-NN method due to the sparsity of the less than 500 data-points in the 128-dimensional feature space and the fact that extracted embeddings lie on an Euclidean space where distances are directly proportional to face similarity.

For a given detected open-mouth face f_i we initialize the confidence vector $c(f_i) = [c_{f_i=1}, c_{f_i=2}, ..., c_{f_i=M}]$ with zeros, where M denotes the number of ground truth classes. We add the the value $1/k$ to the element corresponding to the annotated class of the k^{th} neighbor.

Let $F = \{f_1, f_2, ..., f_N\}$ be the set of N detected open-mouth faces and $c(f_i)$ holds the confidence values $c_{f_i=j}$ of frame f_i belonging to class j. The accumulated confidence $\mathbf{c}_j$ for class j results to

$$\mathbf{c}_j = \sum_{i=1}^{N} c_{f_i=j} \tag{2}$$

and the label l is finally assigned as

$$l = \underset{j}{\operatorname{argmax}}\, \mathbf{c}_j. \tag{3}$$

3. DATA

3.1 Annotated image collection

In the scope of this study we gathered training dataset containing images of flamenco singers. For 10 singers, 3 female and 7 male, we gathered 50 publicly available images each. Images in which no face was detected were discarded, leaving a total of 478 images in the training set.

3.2 Video collection

We gathered a total of 30 videos, 3 videos of each singer in the training database. All videos were taken from online video sharing platforms and apart from the singer at least one more person is seen on stage. The quality ranges from amateur mobile recordings to professional video clip and live performance recording productions. The contained material includes live concerts, private gatherings, excerpts taken from documentaries and music videos.

3.3 Baseline audio collection

In order to compare our approach to state of the art audio-based singer identification methods (Section 4.1), we gathered an additional 10 audio tracks for each singer. The recordings were taken partly from the CorpusCOFLA (Kroher et al., 2016) database and partly from private collections.

4. EXPERIMENTAL EVALUATION

4.1 Baseline method

State of the art audio-based singer identification methods, i.e. Zhang (2003) and Tsai & Lee (2012) follow a common processing framework: A machine learning model is trained on audio descriptors extracted frame-wise from an annotated database. For each frame in the unlabeled audio recording, the same features are extracted and evaluated against the learned model. Finally, the label is assigned based on a majority vote among the frame classifications.

Here, we implemented a baseline approach following this framework. From the annotated recordings in the audio training database, we first extract singing voice segments using an unsupervised method proposed by Pikrakis et al. (2016), which has given reliable results for flamenco recordings. From these segments we then extract the mel-frequency cepstral coefficients (MFCCs) in non-overlapping windows of 50ms length.

As in Zhang (2003), we train a Gaussian mixture model (GMM) for each singer and investigate different values for the number of components C. In the test stage, we extract the same features from the audio track of each unlabeled video, evaluate against the pre-trained GMMs and assign a label based on majority vote over all frames.

4.2 Results

The results of the experimental evaluation by means of correctly classified instances are shown in Table 1. The audio-based baseline method achieves 73.3% classification accuracy among the 10 candidates in the dataset. This is in line with the results reported in Kroher & Gómez (2014) where 86.7% were achieved among 5 candidates. The proposed image-based approach achieves a significantly higher accuracy of 90% for all investigated values for k.

classifier	accuracy
baseline, $C = 2$	66.7%
baseline, $C = 4$	73.3%
baseline, $C = 8$	70.0%
proposed, $k = 1$	90.0%
proposed, $k = 3$	90.0%
proposed, $k = 5$	90.0%

Table 1: Experimental results for audio- and video-based singer identification.

5. CONCLUSIONS

We presented an image-based singer identification system for flamenco videos. We use state of the art image processing techniques to detect faces in video frames and decide based on facial landmarks if a detected face belongs to the singer. Using a learned feature representation, we compare the resulting face images against a dataset and assign a label based on the labels of the nearest neighbors. An experimental evaluation has shown that the method gives promising results compared to an audio-based baseline method and consequently, the video contains valuable information for identifying the singer.

Future work can further explore the potential of image processing for music analysis in various ways: The process of gathering singer images can be automated through the use of web mining techniques. Furthermore, the detected sequences showing the singer's face can be used for emotion recognition or video thumbnail generation. In addition, hybrid approaches to singer identification combining both, audio and video features, could be explored.

6. ACKNOWLEDGEMENTS

This research has received funding from the projects COFLA2 (Junta de Andalucía, P12-TIC-1362), GALGO (Spanish Ministry of Economy and Competitiveness, MTM2016-76272-R AEI/FEDER,UE) and CONNECT (EU-H2020/MSCA under grant agreement 2016-734922).

7. REFERENCES

Ambos, B., Ludwiczuk, B., & Satyanarayanan, M. (2016). Openface: A general-purpose face recognition library with mobile applications. Technical Report CMU-CS-16-118, CMU School of Computer Science.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science*, *10*(3), 243–248.

Cai, W., Li, Q., & Guan, X. (2011). Automatic singer identification based on auditory features. In *Natural Computation (ICNC), 2011 Seventh International Conference on*, volume 3, (pp. 1624–1628). IEEE.

Dalal, N. & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Computer Society Conference on Vision and Pattern Recognition*.

Fix, E. & Hodges Jr, J. L. (1951). Discriminatory analysis-nonparametric discrimination: consistency properties. Technical report, DTIC Document.

Fujihara, H., Goto, M., Kitahara, T., & Okuno, H. G. (2010). A modeling of singing voice robust to accompaniment sounds and its application to singer identification and vocal-timbre-similarity-based music information retrieval. *IEEE Transactions on Audio, Speech, and Language Processing*, *18*(3), 638–648.

Holzapfel, A. & Stylianou, Y. (2007). Singer identification in rembetiko music. *Proc. SMC*, *7*, 23–26.

Jafri, R. & Arabnia, H. R. (2009). A survey of face recognition techniques. *Jips*, *5*(2), 41–68.

Kazemi, V. & Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.*

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, *10*, 1755–1758.

Kroher, N., Díaz-Báñez, J.-M., Mora, J., & Gómez, E. (2016). Corpus cofla: a research corpus for the computational study of flamenco music. *Journal on Computing and Cultural Heritage (JOCCH)*, *9*(2), 10.

Kroher, N. & Gómez, E. (2014). Automatic singer identification for improvisational styles based on vibrato, timbre and statistical performance descriptors. In *Proceedings of the Sound and Music Computing Conference*.

Lagrange, M., Ozerov, A., & Vincent, E. (2012). Robust singer identification in polyphonic music using melody enhancement and uncertainty-based learning. In *13th International Society for Music Information Retrieval Conference (ISMIR).*

Learned-Miller, E., Huang, G. B., RoyChowdhury, A., Li, H., & Hua, G. (2016). Labeled faces in the wild: A survey. In *Advances in Face Detection and Facial Image Analysis* (pp. 189–248). Springer.

Lee, J.-H. (2005). Automatic video management system using face recognition and mpeg-7 visual descriptors. *ETRI journal*, *27*(6), 806–809.

Moon, H. (2004). Biometrics person authentication using projection-based face recognition system in verification scenario. In *Biometric Authentication* (pp. 207–213). Springer.

Ng, H.-W. & Winkler, S. (2014). A data-driven approach to cleaning large face datasets. In *Proceedings of the IEEE International Conference on Image Processing (ICIP).*

Nwe, T. L. & Li, H. (2008). On fusion of timbre-motivated features for singing voice detection and singer identification. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, (pp. 2225–2228). IEEE.

Pikrakis, A., Kopsinis, Y., Kroher, N., & Díaz-Báñez, J.-M. (2016). Unsupervised singing voice detection using dictionary learning. In *Signal Processing Conference (EUSIPCO), 2016 24th European*, (pp. 1212–1216). IEEE.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Shen, J., Shepherd, J., Cui, B., & Tan, K.-L. (2009). A novel framework for efficient automated singer identification in large music databases. *ACM Transactions on Information Systems (TOIS)*, *27*(3), 18.

Sridhar, R. & Geetha, T. (2008). Music information retrieval of carnatic songs based on carnatic music singer identification. In *Computer and Electrical Engineering, 2008. ICCEE 2008. International Conference on*, (pp. 407–411). IEEE.

Tsai, W.-H. & Lee, H.-C. (2012). Automatic singer identification based on speech-derived models. *International Journal of Future Computer and Communication*, *1*(2), 94.

Yi, D., Lei, Z., Liao, S., & Li, S. Z. (2014). Learning face representation from scratch. Technical Report 1411.7923, arXiv prepring.

Zhang, T. (2003). Automatic singer identification. In *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*, volume 1, (pp. I–33). IEEE.

DEMYSTIFYING FLOWS BASED ON TRANSITIONAL PROPERTIES OF IMPROVISATION IN HINDUSTANI MUSIC

Achyuth Narayan Samudrala, Navjyoti Singh
Centre for Exact Humanities
sachyuth.narayan@research.iiit.ac.in, navjyoti@iiit.ac.in

ABSTRACT

This paper explores the relevance/importance of transitions in improvisational decorations to understand the nature of flows in Hindustani music. Hindustani Music is fundamentally improvisational in nature. The artist's main objective is to evoke the mood of the rga (melodic-mode) and this is achieved with the delicate explorations of flourishes and decorations. Music is cognitively associated with these decorations instead of the actual underlying notes. These decorations along with the varied methods of articulating notes together constitute flows. We call them flows due to the movement/change in the frequency/amplitude domains and they are mainly characterized by the rate/nature of change. We show that the sequences of change are fundamental to the idea of flows. We represent them by the first derivative and second derivative of a combination of frequency and amplitude data and then cluster them based on a custom defined distance. We successfully run spectral clustering on a pairwise affinity matrix and achieve an accuracy of 81% in distinguishing between Murki vs Kan and 84.5% in Andolan vs Ghaseet. We thus develop a novel method of content based music analysis which has applications in music search, recommendation, retrieval and pedagogy.

1. INTRODUCTION

In Hindustani music improvisation is of paramount importance. The artist has a lot of freedom with which he renders the performance by using his own style of articulating notes and melodic patterns. Bhatkhande (1934); Bagchee (1998) In fact an artist is also judged by his ability to improvise well. These improvisations are mainly done using various alankaar and alankaran. Mukerji (2014) These can vary from singing a group of notes that express the meaning (bhava) of the raga to the manner of articulation of notes. For instance, execution of notes with a certain amount of shaking or a smooth transition from one note to another while touching other microtones.

Every performance in Hindustani Music is based on a melodic mode (raga). Bagchee (1998) Each raga can evoke a harmonious aesthetic state given that there is a proper rendering of notes and the improvisation is aesthetically sound. Cognitively raagas are associated with these decorations. These along with the ways in which notes are approached, sustained and quitted constitute flows. We call them flows due to the constant change in the frequency and amplitude domains. These flows are a prominent part of a performance and vary from performer to performer. In this paper we want to capture the importance of rate of change in frequency and amplitude domains to conceptualize flows. We do this by clustering known groups of alankaaran (one type of flow) based on a custom distance. This distance compares the sequences of first derivatives and second derivatives of a combination of frequency and amplitude data. The affinity/similarity between two flows is more when there is less distance between the sequences of change. We verify this hypothesis by first creating a pairwise distance matrix, then we use a heat kernel to convert this into an affinity matrix on which we run a spectral clustering algorithm. Since we know the types of flows in the given dataset we then verify if similar flows are being grouped into the correct known labels. We also try to find representative exemplars within these flows by later running the affinity propagation algorithm. The various steps involved in the procedure are shown in figure 1.

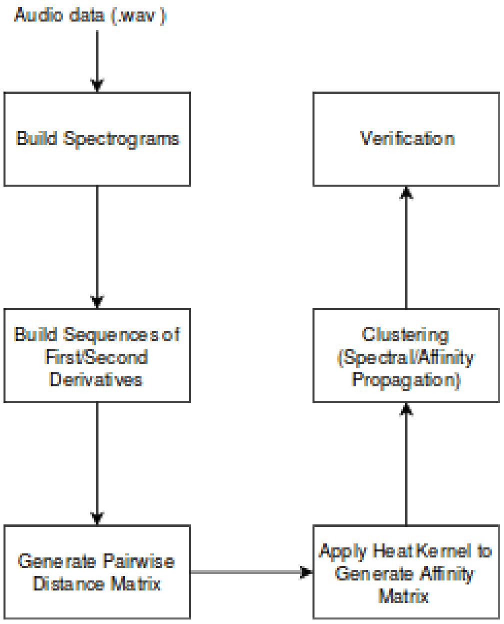

Audio data (.wav)
Build Spectrograms
Build Sequences of First/Second Derivatives
Generate Pairwise Distance Matrix
Apply Heat Kernel to Generate Affinity Matrix
Clustering (Spectral/Affinity Propagation)
Verification

Figure 1: Flowchart of the complete procedure.

2. BACKGROUND

Alankaaran (an act of Ornamentation) is constituted by alankaar, gamak, sthaya and kaaku. Mukerji (2014) A specific group of notes that get its meaning in context of a musical phrase are known as alankaar. Gamak is the execution of notes with a certain amount of shaking. Sthaya is a combination of a certain notes of the raga which expresses a certain emotion of the raga. Lastly, Kaaku refers to the manner of articulating a particular note. Mukerji (2014)

ISBN: 978-84-697-2303-6

These are the different modes in which Alankaaran can be achieved. We are interested in those modes of alankaaran which involve fair bit of improvisation and are specific to the performer/music style. Alankaar further consists of Varnaalankaar, Chhandalankaar, Varnatiriktaalankaar and Vadanbheda. Varnaalankaar are alankaars using groups of notes. When note-groups are used in rhythmic patterns Chhandalankaar arises. When alankaaran is achieved by creating variations in the articulation of notes, varnatiriktalankaar is achieved. Mukerji (2014) In this paper we mainly focus on these kinds of alankaars. We particularly focus on Murki, Kan, Andolan, Meends and Ghaseet.

1. Murki: A murki is cluster of notes that sounds like a short, subtle taan. It is a fast and delicate ornamentation employing two or more notes. Mukerji (2014)

2. Kan: Kan are the linking or grace notes. They are played in a very subtle manner. Mukerji (2014)

3. Andolan: The Andolan alankar is a gentle swing or oscillation that starts from a fixed note and touches the periphery of an adjacent note. Mukerji (2014)

4. Meend: Meend refers to the glissando or a glide in the wave from one note to another. Mukerji (2014)

5. Ghaseet: This is a kind of meend specific to stringed instruments. While its literal meaning translates to pull, it usually refers to a fast glide. Mukerji (2014)

North Indian classical music is not a staccato form and the transitions between notes are of utmost importance. The variations brought about in the rendering of notes such as: 1.) gentle oscillations, 2.) subtle increase/decrease in volume/speed, 3.) sudden variation in volume/speed, 4.) sustenance on one particular note for long, 5.) touching micro-tonal shrutis whilst performing alankaaran etc. are instrumental to the performance and are representative of the style of music. Cognitively we associate a particular performance with these improvisations and methods of articulation. More specifically, we associate music with such melodic patterns. These melodic patterns involving different methods of articulation of notes, transitions between notes and improvisational motifs constitute flows. Flows are characterised by rate of change. For instance, in a ghaseet there is a rapid change in frequency in a short interval of time since multiple micro-tonal shrutis are touched sequentially. Mukerji (2014) Whereas, in a krintan the change is more subtle. In mathematical terms flows can be understood by looking at the derivative of the signal in frequency/amplitude domains. In this paper we try to capture the change in frequency and amplitude domain together since there is a constant change in both the frequency and the amplitude domain when a musician is trying to achieve alankaaran. For instance, when andolan is being performed, the musician gently touches the periphery of other notes. Thus along with a change in the frequency domain there is also a notable change in the amplitude. Hence, we look at the first/second derivative of amplitude weighted frequency data.

3. RELATED WORK

Most of the earlier work done in the field of ornament analysis use some or the other variant of pitch tracking. In Pratyush (2010), Pratyush uses time series matching to find the distance between two ornaments which are represented by its pitch sequence. The results shown here are highly questionable because the dataset is smaller than a hundred samples and it has very little variety. In S. S. Miryala & Choudhury (2013), an algorithm is proposed for automatic transcription, using line fitting to obtain canonical representation of pitch curves in terms of straight lines and points of inflection and then using template matching to identify vocal expressions. However, the algorithm often fails to distinguish between steady notes and glides. Also, the templates used are not exhaustive, they dont cover a lot of variety of ornaments. Gupta & Rao. (2012) proposed objective methods to assess the quality of ornamentation in Indian music performed by a singer. They take into account an ideal singer as a model or reference and compare reference meends with the meends sung by different singers on the basis of (i) point to point error calculation; (ii) Dynamic time warping and (iii) polynomial curve fit based matching.

Narayan & Singh (2014) use iterative template matching to detect ornaments in Dhrupad, a variant of North Indian Classical Music. This study again is heavily dependent on the templates being used. The intrinsic rate of change properties of ornaments are not being captured. Time warping in general has been the distance measure used to compare time dependent sequences. Gulati et al. (2016); Joe Cheri Ross & Rao (2012) In Narayan & Singh (2015) they build on a model to study the consonance of notes being used in various alankaar. Joe Cheri Ross & Rao (2012) propose various similarity metrics to detect melodic motifs in Hindustani music. Along with the conventional DTW they also use piecewise aggregate approximation to convert a non uniform length time series to a uniform length dimension reduced sequence of symbols. A given time series is aggregated into uniform W length sequences and Euclidean distance is used as the similarity measure.

Gulati (2016), develops computational approaches for analyzing high-level melodic aspects of music performances in Indian art music. For extracting melodic features he uses the Melodia pitch tracking algorithm and ignores the energy and velocity information. Gulati (2016) ignores the importance of these factors in representation of melodic patterns or ornaments. He then uses normalized and error corrected pitch tracked information along with DTW for mining patterns. To identify musically meaningful patterns, he exploits the relationships between the discovered patterns by performing a network analysis, exploiting the topological properties of the network. Gulati (2016) recognises that these patterns are the building blocks of melodic structures in both improvisation and composition. Thus, that they are fundamental to the description of audio collections. He also recognizes that it is important to identify these patterns for interacting with large volumes of audio recordings, and for developing novel tools to facilitate music pedagogy.

The true nature of ornaments is that they are specific to the performer and vary a lot from performer to performer and style to style. The nature of rate of change cannot be captured just by looking at the pitch tracking information, not just because it ignores the energy changes but also because there might be subtler changes in the partials/overtones of fundamental frequency which might hold a lot of importance in understanding the rate of change in ornament based and other flows. This is the biggest difference between the approach in this paper and any of the recent work done in the field of ornament analysis.

4. FEATURE EXTRACTION

In this section we will discuss about the various steps involved in representing flows in mathematical terms. As discussed earlier a flow can be mathematically expressed as the first/second derivative of the frequency/amplitude data. In this paper we mainly are interested in certain ornaments for the study of flows. We have recorded various ornaments played on sitar and stored them as separate ornament flows. Now that the flows have been recorded we initially compute their spectrograms. Spectrograms can be used as a way of visualizing the change of a non-stationary signals frequency content over time. Since we are interested in looking at the changes at a sub-second level we look at the frequency content at every hundred millisecond window. The audio files are sampled at 44100 Hz, so we consider the length of each segment to be 4410 samples.

After the spectrograms have been created, we are interested in finding the first and second derivatives of frequency and amplitude content. We approximate the calculation of derivatives by computing delta and double deltas. As discussed earlier, during the act of ornamentation there are changes both in frequency and amplitude content at once. To capture this dependence of frequency and amplitude while computing the derivatives, we dont look at frequency and amplitude separately but take a product of frequency and amplitude. Then we compute the derivatives of this combined measure. We have thus chosen this measure over a linear combination of frequency and amplitude content. To make the idea clearer lets take an example of a flow X, whose spectrogram has been computed. X has frequencies in the range (Fmin, Fmax) and in each 100 ms time window these frequencies can have varying energy content ranging from (0, Emax). Now lets say in the first window F has the highest energy content E. We compute sqrt(+E)*F and call this combined measure C which stores the product for the highest energy frequency in the first window. Now we go to the next time window and compute C by looking at the maximum energy frequency and amplitude. Then D1 is the difference between C in time window 2 and 1. Similarly computing for all time windows we get sequences of deltas for the highest energy frequency (D1, D2). If there are n time instances then we have the delta sequence of length n-1. These sequences are then calculated for the next highest power frequencies. At the end of this, each flow will have n sequences of deltas wherein n is the total number of frequency components in the spectrogram. Similarly we compute the second derivatives from the first derivative data. For an n length first delta sequence we get a n-1 length second derivative sequence. Here again we have n second derivative sequences for n number of frequency components in the spectrogram of the flow being considered. In conclusion, a flow which has n frequency components and t time windows can be represented by two sequences: n first derivative sequences and n second derivative sequences, of length t-1 and t-2 respectively.

5. DISTANCE MEASURE

Now that we have sequences of change representing the flows given, we come up with a distance metric to identify the similarity between two flows. Fundamentally to compare two time sequences the most commonly used idea is that of Dynamic Time Warping. Dynamic Time Warping is a dynamic programming algorithm which tries to find an optimal alignment between two time domain sequences. Mller (2007) Since we are interested in capturing the similarity between sequences of change this is suitable for our problem. Consider two time domain sequences X = [x1, x2...xM] and Y = [y1, y2...yN]. Evaluating the local cost between each pair of X and Y we obtain a distance matrix D(X, Y). Using this matrix we try to find the optimal alignment using the recursive formula:

$$D(i,j) = d(i,j) + min(D(i,j-1), \\ D(i-1,j), D(i-1,j-1)) \quad (1)$$

Here d(i,j) is the local cost measure between Xi and Yj. We use Euclidean distance as our local cost measure. D(M,N) now stores the value of the optimal alignment. Now that we have established a metric to find the distance between two time varying sequences lets integrate it into our setting of multiple sequence representation.

Lets consider that we have two flows X and Y. X has n1 sequences of first derivative data. Y has n2 sequences of first derivative data. The difference is number of sequences is due to the varying frequency content in both the sequences. Thus we only look at top k sequences which correspond to sequences of top k highest energy frequencies in the spectrogram. So for top K highest power frequencies the first derivative distance between two flows can be given by:

$$FDdist = EuclideanNorm(1st - distance, \\ ..ith - distance, ..kth - distance) \quad (2)$$

Wherein, to compute the ith-distance we compare the ith sequence of X with ith sequence of Y using the earlier defined Dynamic Time Warping algorithm. Similarly the second derivative distance between two flows can be given by:

$$SDdist = EuclideanNorm(1st - distance, \\ ..ith - distance, ..kth - distance) \quad (3)$$

Where to compute the ith-distance we compare the ith second derivative sequence of X with ith second derivative sequence of Y. The total distance between two flows can then be given as:

$$Dist(X, Y) = FDdist + SDdist \quad (4)$$

Where Dist(X,Y) is the distance between two flows X and Y.

6. CLUSTERING

Since the distance metric between two flows has been established we can now perform clustering to check how well we can group the flows into their known labels. If we have n number of flows in our dataset we construct a nxn matrix of pairwise distances. We had initially run the K-Medoids clustering on this distance matrix to find that the clustering on distance wasnt very good. Kaufman & Rousseeuw (1987) Therefore now we construct an affinity matrix out of the distance matrix using a heat kernel, on which we can run the spectral clustering algorithm. In the context of clustering, an affinity measure is just the converse of a distance i.e. a distance of 0 means highest affinity. If values in the affinity matrix are not well distributed the spectral problem will be singular and not solvable. Thus we apply the below heat kernel on the given distance matrix:

$$similarity = np.exp(\frac{-beta * distance}{distance.std()}) \quad (5)$$

where np is the numpy library in python. It can be approximated as:

$$similarity = np.exp(\frac{-distance^2}{(2. * (distance.max() - distance.min())^2)}) \quad (6)$$

6.1 Spectral Clustering

The goal of spectral clustering is to cluster data that is connected but not necessarily compact or clustered within convex boundaries. Ng et al. (2001) It is efficient if the affinity matrix is sparse. It needs us to specify the number of clusters upfront and works well for small number of clusters. For two clusters, it solves a convex relaxation of the normalised cuts problem on the similarity graph. In scikit-learn spectral clustering does a low-dimension embedding of the affinity matrix between samples, followed by a K-Means in the low dimensional space. Steps involved in this type of clustering:

1. First we construct an affinity matrix A.
2. Then we construct the graph Laplacian from A. There are many ways to define a Laplacian. Normalized, generalised, relaxed etc.
3. Compute eigenvalues and eigenvectors of the matrix. Each eigenvector provides information about the connectivity of the graph. The idea of spectral clustering is to cluster the points using these ”k” eigenvectors as features.
4. Map each point to a lower dimensional representation based on the computed eigenvectors.
5. Assign points to clusters based on the new representation.

In this we essentially try to find a transformation of our original space so that we can better represent manifold distances for some manifold that the data is assumed to lie on. Ng et al. (2001) When the data is projected into a lower dimensional space it makes the data easily separable and thus the clustering algorithm works. However, this still retains many properties of K-means since after we find a low dimensional embedding, we run the K-means algorithm. But the fact that we should know the labels before hand is not a problem for us since we have a labeled data-set. However within these labeled flows as well, there might be other intra clusters which we arent aware of because of the subtle nature of variations present in the flows. Thus we also try to find these smaller groups and representative exemplars by running the affinity propagation algorithm. Dueck (2007) Spectral clustering works wells for us also because the size of the data-set is small enough. If there were items in our data-set in the order of 10^5 then we would have to construct an affinity matrix of size 10^{10} which would bloat up the main memory.

6.2 Affinity Propagation

Affinity propagation is a clustering algorithm which doesn’t need any predefined number of clusters to be given as input. It finds the exemplars which are representative of the clusters in the data-set. Dueck (2007) It views each data point as a node in a network, and recursively transmits real-valued messages along edges of the network. This is done until a good set of exemplars and corresponding clusters emerges. These messages are updated on the basis of formulas that search for minima of a chosen energy function. The magnitude of each message reflects the current affinity that one data point has for choosing another data point as its exemplar.

This method takes as input a real number s(k, k) for each data point k so that data points with larger values of s(k, k) are more likely to be chosen as exemplars. The number of identified clusters is influenced by the values of the input preferences. It is also affected by the message-passing procedure. The algorithm proceeds by alternating two message passing steps, to update the responsibility and the availability matrices. The responsibility matrix has values r(i, k) that reflect how well-suited xk is to serve as the exemplar for xi, relative to other candidate exemplars for xi. The availability matrix has values a(i, k) that reflects accumulated evidence as to how well-suited it would be for xi to pick xk as exemplar, taking into account the support from other points preference for xk as an exemplar. These matrices can be seen as log probability ratios.

Initially the availability is zero and then responsibilities are updated by the rule:

$$r(i,k) \leftarrow s(i,k) - max\{a(i,k') + s(i,k')\}, \quad where \quad k' \quad s.t. \quad k' \neq k \tag{7}$$

Then the availability is updated as follows:

$$a(i,k) \leftarrow min\{0, r(k,k) + \sum_{i' \ni \{i,k\}} max\{0, r(i',k)\}\} for \quad i \neq k \quad and \tag{8}$$

$$a(k,k) = \sum_{i' \neq k} max(0, r(i',k)) \tag{9}$$

The iterations are performed until convergence, at which point the final exemplars are chosen, and hence the final clustering is given. Dueck (2007) The data points whose responsibility+availability is positive are chosen as exemplars.

7. RESULTS

Our data-set consists of few hundreds of samples of ornament flows comprising of Murki, Kan, Andolan, Ghaseet (type of meend). These ornaments were performed in different ragas to create a comprehensive data-set. The ornaments were performed in ragas: Malhar, Marwa, Mishra Piloo, Bhagyashree and Yaman. We test the validity of the distance metric and our representation by first performing spectral clustering and then running the affinity propagation algorithm. We first look at accuracy as the performance evaluation metric. Accuracy is given by:

$$Accuracy = \frac{tp + tn}{tp + fp + tn + fn} \tag{10}$$

where tp = true positive, tn = true negative, fp = false positive, fn = false negative We find that the accuracy of clustering Murki and Kan by giving a predefined k equal to 2 is 0.81, i.e, the clustering algorithm correctly clustered murki and kan into their respective groups in 81% of the cases. The accuracy of clustering Andolan and Ghaseet in a similar fashion is 0.84. These accuracies are very high considering that in our data-set of ornaments we had a variety not just in terms of ragas but also in the articulation of these ornaments. For instance, our kan data-set has both two note and three note variants. The length of the ornaments in our data-set are also varied ranging from one second to many seconds. This shows the suitability of our distance metric and representation for varying length ornaments. We compare Murki/Kan and Andolan/Ghaseet because they are fairly different and they usually dont occur together. In general musicians tend to combine andolan and murki in performance. Even ghaseet co-occurs with murki and andolan sometimes.

The Fowlkes-Mallows score FMI is defined as the geometric mean of the pairwise precision and recall. Fowkles & Mallows (1983) It ranges from 0 to 1 and high score indicates good similarity between clusters. Perfect labeling has a score of 1.0. FMI is given by:

$$FMI = \frac{TP}{\sqrt{(TP + FP)(TP + FN)}} \tag{11}$$

where TP = True Positives, FP = False Positives, FN = False Negatives

FMI scores of 0.70 and 0.77 indicate that the precision and recall of our clustering is also very good. The values corresponding to performance evaluation metrics shown in Table 1 prove that the representation and distance metric have performed well in grouping similar flows together.

However we also test the legitimacy of our clustering by other metrics as well. Adjusted rand index (ARI) Hubert & Arabie (1985) is the number of pairs of objects that are either in the same group or in different groups in both partitions divided by the total number of pairs of objects. It ranges from -1 to 1 and a positive score indicates similar clustering.

$$ARI = \frac{RI - E[RI]}{max(RI) - E[RI]} \tag{12}$$

where RI is the rand index. The raw rand index is given by:

$$RI = \frac{a + b}{C} \tag{13}$$

where a = the number of pairs of elements that are in the same set in ground truth and in actual labels, b = the number of pairs of elements that are not in the same set in ground truth and in actual labels, C = is the total number of possible pairs in the dataset.

The ARI is positive and reasonably high for both clustering experiments. Higher ARI shows that the partitions are in agreement with each other. ARI is larger for the case of Andolan vs Ghaseet.

The Mutual Information Strehl & Ghosh (2002) is a function that measures the agreement of the two assignments, ignoring permutations. Bad/Independent labellings have negative scores. The calculated value of the mutual information is not adjusted for chance and will tend to increase as the number of different labels (clusters) increases, regardless of the actual amount of mutual information between the label assignments. Given that we just have two labels this is not a very good metric to assess our results. However a positive score is a good indicator.

Murki vs Kan	**Accuracy**	0.81
	Fowlkes Mallows Score	0.70
	Adjusted Rand Index	0.66
	Adjusted Mutual Information Score	0.46
	Homogeneity	0.47
	Completeness	0.48
	V-measure	0.47
Andolan vs Ghaseet	**Accuracy**	0.84
	Fowlkes Mallows Score	0.77
	Adjusted Rand Index	0.75
	Adjusted Mutual Information Score	0.51
	Homogeneity	0.52
	Completeness	0.54
	V-measure	0.53

Table 1: Table containing values corresponding to various clustering metrics to validate the quality of spectral clustering.

Given that we have the knowledge of the ground truth class assignments, it is possible to define some intuitive metric using conditional entropy analysis. Homogeneity tells whether each cluster contains only members of a single class, while Completeness tells whether all members of a given class are assigned to the same cluster. Rosenberg & Hirschberg (2007) We observe that the homogeneity and completeness scores are low, which implies that there is a little bit of overlap in clustering. But since these functions increase at the pace of the logarithmic function, the numbers might improve significantly with more samples or clusters. V-measure or the harmonic mean is actually equivalent to the mutual information normalized by the sum of the label entropies. Rosenberg & Hirschberg (2007) Previously an analysis has been done on the impact of the number of clusters and number of samples on these metrics dependent on conditional entropy. scikit-learn developers (scikit-learn developers) The observation is that the values of these metrics increase significantly with increase in number of samples and the number of clusters. Since our dataset is quite small these metrics can be ignored. For small sample sizes or number of clusters it is safer to use an adjusted index such as the Adjusted Rand Index (ARI).

Murki vs Kan	**Number of Clusters**	15
	Silhouette Coefficient	0.65
Andolan vs Ghaseet	**Number of Clusters**	13
	Silhouette Coefficient	0.75

Table 2: Table containing the values corresponding to metrics of affinity propagation.

To verify the results of the affinity propagation algorithm we look at the silhouette coefficient. Rousseeuw (1987) Higher the value implies better the clustering assignments. The Silhouette Coefficient Rousseeuw (1987) is defined for each sample and is composed of two scores a: The mean distance between a sample and all other points in the same class. and b: The mean distance between a sample and all other points in the next nearest cluster. The Silhouette Coefficient s for a single sample is then given as:

$$s = \frac{(b-a)}{max(a,b)} \tag{14}$$

The exemplars (centers) extracted by this algorithm have been verified and have proven to show distinct musical significance.

8. CONCLUSIONS

We have shown that the property of rate of change is fundamental to the idea of flows and can be represented by the first derivative and second derivative data. We were able to achieve very high clustering accuracies by using this representation and our custom defined distance measure. We achieved an accuracy of 81% in distinguishing between Murki vs Kan and 84.5% in Andolan vs Ghaseet. This again shows that transition is at the heart of the definition of flows and ornaments. However we have not compared some other flows such as the approaching of notes/quitting of notes separately across performances. Further study could take up only looking at patterns between these types of flows across individual performers and gharanas. Since this dataset is still very small, a comprehensive dataset can be built to check for the exhaustiveness of this hypothesis.

In this paper we have just looked at one combination of frequency and amplitude data. Many other combinations can be explored to find the correct dependence. This can further lead to different representations of flows.

This theory of rate of change might be suitable for any other form of monophonic music as well where melody plays a major role. The transitional nature of melodic phrases can be expressed using a similar representation based out first and second derivative data. This type of content based audio analysis plays a prominent role in interacting with large volumes of audio recordings and also for developing novel tools to facilitate music pedagogy. It can have applications in music information retrieval, search and recommendation systems as well.

9. REFERENCES

Bagchee, S. (1998). Nad: understanding raga music. *Eeshwar - illustrated edition*.

Bhatkhande, V. N. (1934). Hindusthani sangeet paddhati. *Sangeet Karyalaya*.

Dueck, B. J. F. D. (2007). Clustering by passing messages between data points. *Science*, *315*, 972976.

Fowkles, E. B. & Mallows, C. L. (1983). A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association.*

Gulati, S. (2016). *Computational Approaches for Melodic Description in Indian Art Music Corpora*. PhD thesis, Universitat Pompeu Fabra, Barcelona.

Gulati, S., Serrà, J., Ishwar, V., & Serra, X. (2016). Discovering rāga motifs by characterizing communities in networks of melodic patterns. In *41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016)*, (pp. 286–290)., Shanghai, China. IEEE, IEEE.

Gupta, C. & Rao., P. (2012). Objective assessment of ornamentation in indian classical singing. In *CMMR'11 Proceedings of the 8th international conference on Speech, Sound and Music Processing: embracing research in India*, (pp. 1–25)., Berlin, Heidelberg. Springer-Verlag.

Hubert, L. & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2: 193*.

Joe Cheri Ross, V. T. & Rao, P. (2012). Detecting melodic motifs from audio for hindustani classical music. In *Proc. 13th International Society for Music Information Retrieval Conference*, Porto.

Kaufman, L. & Rousseeuw, P. (1987). Clustering by means of medoids. *Statistical Data Analysis Based on the L 1 Norm and Related Methods*, 405416.

Mukerji, B. (2014). *An analytical study of improvisation in hindustani classical music*. PhD thesis, University of Delhi.

Mller, M. (2007). *Information Retrieval for Music and Motion*. Berlin, Heidelberg: Springer.

Narayan, A. & Singh, N. (2014). Detection of micro-tonal ornamentations in dhrupad using a dynamic programming approach. In *Proc. 9th Conference on Interdisciplinary Musicology CIM14*, Berlin.

Narayan, A. & Singh, N. (2015). Consonance of micro-tonal ornamentation in melodic contexts. In *Proc. of the 11th International Symposium on CMMR*, Plymouth, UK.

Ng, A. Y., Jordan, M. I., & Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. In *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, (pp. 849–856). MIT Press.

Pratyush (2010). Analysis and classification of ornaments in north indian (hindustani) classical music. Master's thesis, Universitat Pompeu Fabra.

Rosenberg, A. & Hirschberg, J. (2007). V-measure: A conditional entropy-based external cluster evaluation measure. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, (pp. 410–420)., Prague, Czech Republic. Association for Computational Linguistics.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Computational and Applied Mathematics*, *20*, 53–65.

S. S. Miryala, K. Bali, R. B. & Choudhury, M. (2013). Automatically identifying vocal expressions for music transcription. In *Proc. ISMIR*, (pp. 239–244)., Brazil.

scikit-learn developers. Adjustment for chance in clustering performance evaluation.

Strehl, A. & Ghosh, J. (2002). Cluster ensembles a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, *3*, 583–617.

Oral session 4

Folk Music Analysis
MÁLAGA, Spain
14-16 June 2017

SOURCE SEPARATION BY SCORE SYNTHESIS IN SPANISH FOLK MUSIC

Sergio P. Paqué, Marcelo Caetano, José L. Santacruz,
Lorenzo J. Tardón, Isabel Barbancho

ATIC Research Group, ETSI Telecomunicación,
Universidad de Málaga, Andalucía Tech, Málaga, Spain
sergioppaque@ic.uma.es, mcaetano@ic.uma.es, jls@ic.uma.es,
lorenzo@ic.uma.es, ibp@ic.uma.es

ABSTRACT

The aim of source separation of polyphonic music is to obtain one track per musical instrument. Source separation of musical performances is a notoriously difficult problem because the instruments commonly overlap both in time and frequency. Spanish folk music can be very challenging to separate due to the large number of ornaments, musical flourishes, and dynamics. Additionally, each performance also can be different from the general score depending on expressivity, interpretation, arrangement, among other factors. In this paper, we propose to use prior information from the score to improve the separation of recordings of Spanish folklore music into isolated musical instruments, each playing a different voice. We use the MIDI score to synthesize a first approximation of the musical performance. Then, the synthesized data is aligned with the recording of the performance using Dynamic Time Warping (DTW). Finally, the time-aligned approximation is used as the prior model in a Probabilistic Latent Component Analysis (PLCA), which separates the recording into different instrumental tracks. This model provides us a good prior reconstruction of the real instrument, as well as a fit of the onsets and musical ornaments.

1. INTRODUCTION

Source Separation has always been a major challenge in constant improvement. In various applications (as remixing, sampling or academic needs), isolated instruments or sounds from recorded mixtures have been required. Especially in Spanish Folk Music, playing with an entire band is more appropiate for practicing, so, using voices from real recordings is useful for musical education. MIDI scores are now a very abundant material, easy to process and from which we can extract a lot of general information about a musical piece, such as the pitch, intensity or duration of each note, as well as the different voices that perform each instrument.

Therefore, although there are methods that start from the premise of not knowing information of the piece called blind methods [1, 2, 4, 11, 12], many of the methods of separation of sources make use of this previous information to guide the separation. Every performance of the same musical piece varies from the score in terms of dynamics, tempo and notes duration. Among other factors, these microvariations in performance can be attributed to expressiveness. It has been a researching object in Music Information Retrieval last years to correct the music representation, having a representation that correspond exactly with the recording. In [6, 7], the authors propose using Dynamic Time Warping (DTW) to align polyphonic music to scores using a measure called Peak Structure Distance (PSD), which is derived from the spectrum of audio and from synthetic spectra computed from score data. In recent years, a growing number of source separation methods based on spectral decomposition have been proposed. Prior information obtained from the score could be used in many forms. In Soundprism [14], the score guides the creation of harmonic filters for each source. One of the most used methods in spectrogram decomposition is the Non-Negative Matrix Factorization, that provides time activations and frequency components that are used to reconstruct the separate signals [9, 10]. NMF is also used as a Blind Source Separation method [1, 4]. The Probabilistic Latent Component Analysis (PLCA) is another factorization method, equivalent to the NMF, but with a probabilistic iterative vision of the computation of time-frequency components [3, 5, 8, 13]. The components are modeled to the real recording using an Expectation-Maximization algorithm.

In this work, we propose to use synthesized tracks from the MIDI score as priors to build time activations and frequency spectra for each voice in a PLCA model. Every component of every track will be adapted to the real mixture, so we know upon convergence which components belong to which reconstructed instrument. Soloist instruments in Spanish Music as Piano or Classical Guitar have fast notes and strumming chords. Generating a good prior timbral approximation of the instruments, we will compute the reconstructed instruments easily. Usually, the melodies in Spanish Folk Music rely heavily on musical flourishes such as vibrato, shakes, *grupetto*, etc. and performance variations resulting from expressiveness or interpretation, for example. Thus, PLCA allows to adapt the ideal prior synthesis to the real sound, adjusting the time-frequency variations present in the performance that are not indicated in the score.

The paper is organized as follows. In section 3 we describe the PLCA algorithm proposed in a general form by [8]. Section 4 explain how we use the score and the synthesis in the process of separation, and in section 5 we ex-

ISBN: 978-84-697-2303-6

plain the final point of source extraction. Finally, we describe our particular example in section 6, and use the method to extract a flamenco guitar from a real recording, explaining possible improvements and future in section 7.

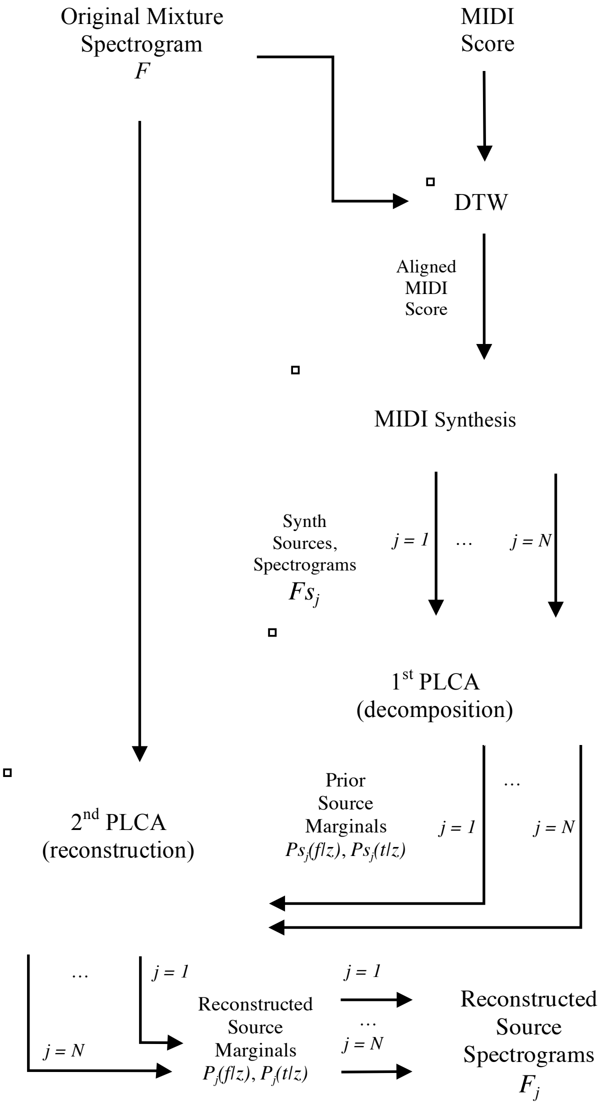

Original Mixture Spectrogram F
MIDI Score
DTW
Aligned MIDI Score
MIDI Synthesis
Synth Sources, Spectrograms Fs_j
j = 1 ... j = N
1st PLCA (decomposition)
2nd PLCA (reconstruction)
Prior Source Marginals Ps_j(f|z), Ps_j(t|z)
j = 1 ... j = N
... j = 1
j = N
Reconstructed Source Marginals P_j(f|z), P_j(t|z)
j = 1 ... j = N
Reconstructed Source Spectrograms F_j

Figure 1. An Overview of the method.

2. OVERVIEW OF THE METHOD

An overview of the complete method is shown in Fig. 1. First, the MIDI score is time-aligned to an audio rendition of the piece using DTW. Then, we compute its spectrogram (Fs_j), and we decompose it using a random-initialized PLCA (1st PLCA), obtaining a combination of 1-dimension marginal distributions in time and frequency, that are modeled to the real spectrogram (F) initializing a 2nd PLCA [3]. The reconstructed source spectrogram is a result of the product a sum of the final components.

3. PLCA

The Probabilistic Latent Component Analysis is an iterative method that allows the decomposition of an N-dimensional distribution of the random variable $X = \{x_1, x_2, ..., x_N\}$. The result of this probabilistic model is a mixture of non-negative marginal distribution (1 dimension), whose products recompose the N-dimensional distribution [5, 8].

3.1. 2D-PLCA General Model

A spectrogram from audio data can be considered a histogram (2-dimensional) distribution over time and frequency. So, the real spectrogram could be analyzed as a non-normalized probabilistic distribution. We separate the 2-dimensional spectrogram into time marginal distributions and frequency marginal distributions, and the original 2-dimensional distribution is the product between the components of each dimension.

So, in audio and music terms, we can say that each marginal distribution is a part of the complete piece. For example, a frequency distribution could correspond to a unitary musical note, and the corresponding time distribution represents the activation and dynamic of the note.

Then, the reconstruction of the decomposed spectrogram:

$$F = \sum_{z}^{K} P(z)P(f|z)P(t|z) \qquad (1)$$

Where F is the reconstructed spectrogram, z is a latent variable (each time-frequency component), K is the number of marginal distribution in each dimension and $P(f|z)$ and $P(t|z)$ are the z marginal 1-D distribution in freq./time. $P(z)$ only contains the weight of the product z in the total mixture. We have a $P(f|z)$ and $P(t|z)$ product for each variable z, see Fig. 2.

We perform an iterative Expectation-Maximization algorithm to compute the marginal distributions and obtain a converging result. The prior marginal distributions could be pre-trained marginals, or we can decompose from zero additional information by initializing random distributions (uniform probability).

So, in each iteration:

- Expectation: We compute the contribution of each the component z (from first to K) to the real mixture $F_{f,t}$.

$$R(z|f,t) = \frac{P(z)P(f|z)P(t|z)}{\sum_{z\prime}^{K} P(z\prime)P(f|z\prime)P(t|z\prime)} \qquad (2)$$

- Maximization: We use this contribution to re-estimate the new marginal distributions, that are closer to the real original spectrogram. We have a first step (components denoted by (*)). When the weights *P(z)* are computed, we calculate de definitive *f/t* marginals.

$$P^*(f|z) = \sum_t (F_p \cdot R(z|f,t)) \quad (3)$$

$$P^*(t|z) = \sum_f (F_p \cdot R(z|f,t)) \quad (4)$$

$$P(z) = \sum_t \sum_f (F_p \cdot R(z|f,t)) \quad (5)$$

where F_p is the reconstructed Spectrogram from the last iteration data product, as in (1).

And, from $P^*(x_i|z)$:

$$P(x_i|z) = \frac{P^*(x_i|z)}{P(z)} \quad (6)$$

Finally, each marginal $P(x_i|z)$ and *P(z)* is normalized to correspond to a probability distribution.

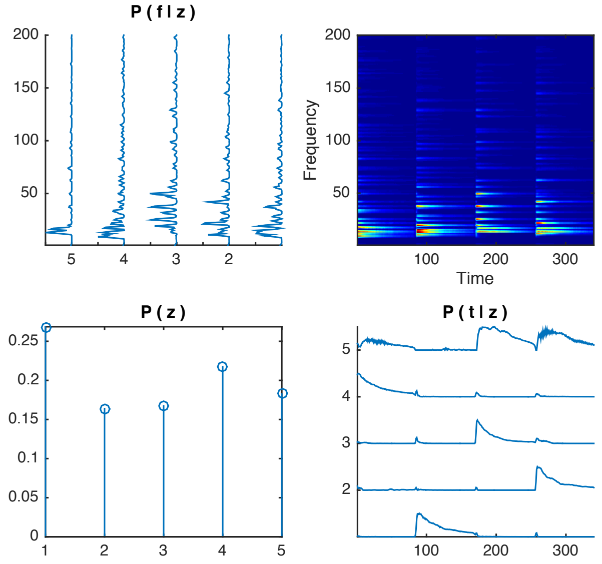


Figure 2. Example of frequency (P(f|z), frequency in y-axis) and time (P(t|z), time in x-axis) components (with respective weights) for a piano chords spectrogram.

4. SCORE ROLE IN SOURCE SEPARATION

4.1. Alignment

It is worth noting that, in Spanish Folk Music, is more frequent to find variations between a written score and a real performance recording, depending on the interpreter and the performance's expressiveness.

The MIDI score (a symbolic representation of the musical piece) allows to know which notes an instrument should be playing at any moment, and it contains ideal information about the tempo, note onsets, durations and notes loudness.

This MIDI data is ideal, opposite to the recording mixture, that probably has different tempo and micro-variations. If we don't synchronize the MIDI info and the original recording, the MIDI data is incorrect and the separation probably doesn't work.

The MIDI score can be synchronized with the audio using a Dynamic Time Warping [6, 7]. With this method, we compare the two sequences (MIDI and audio), and we find the correspondence between them, allowing us to adapt the MIDI. So, the MIDI info will correspond exactly to the original audio, having more accuracy in the process of extraction.

4.2. MIDI Synthesis

The synchronized MIDI score can be used as a reliable guide of what is playing in each moment of the musical piece, and this information allows to locate the source in the spectrogram both in time and frequency.

In this work, we synthesize the separate instruments from the MIDI to have a prior base of the sounds that we separate [3]. Then, we complete this separate sounds with information of the real recording. This synthesized instrument works as a prior frequency/time structure of the individual source, and it's similar to the real sound. It is not a natural sound, but it is a reliable structure that have information about where is placed the instrument that we want to extract in the spectrogram.

One of the most important considerations when we synthesize the instruments from the score is the similarity with the original sound. If the synthesized sound has the correct partials and a good time transient approximation (a good global similarity to the real instrument sound), the PLCA model works better, and we will have a good converging solution, as is explained in [3]. For example, if we try to extract a piano sound (that has even and odd partials) with a synthesized clarinet sound (a frequency spectrum with prominent odd partials), the extracted source will be closer to a clarinet than a piano, because the prior spectral base is not correct and the even partials of the piano are not seen, so, the extraction is wrong.

4.3. Prior Source Structure

The next step of the method is to use the 1st PLCA algorithm to decompose the synthesized audio tracks, so that we go from having two-dimensional structures (frequency/time) to a set of one-dimensional distributions, whose combination is each synthesized source Fs_j.

First, once we have calculated the spectrogram of each musical line of the synthesized instruments that we are going to separate (which we call $Fs_{j,}$ where j is each of the N instruments that are to be extracted), we apply the PLCA algorithm to decompose these spectrograms into time and frequency components ($Ps_j(f|z)$ and $Ps_j(t|z)$).

So, we can have previous components of each source Fs_j, which will then be approximated together to the original mix spectrogram (F) with a second PLCA, while still identifying which components correspond to each source.

5. FINAL 2D-PLCA INDIVIDUAL RECONSTRUCTION

The second time we use the 2nd PLCA algorithm we apply it in a different way. In this case, we initialize the analysis with the components $Ps_j(f|z)$ and $Ps_j(t|z)$ as previous data, trying to decompose the F spectrogram of the original recording.

The iterations in this case start from the marginal distributions from synthesized sounds $Ps_j(f|z)$ and $Ps_j(t|z)$. These components are adapted to the real sound at every iteration, as is shown in Fig. 3.

Finally, we get a combination of marginals $P_j(f|z)$ and $P_j(t|z)$, result of the decomposition of the mixture F, so we can identify which component corresponds to each instrument (j), and reconstruct each spectrogram (reconstructed and extracted spectrograms F_j from sources $j = 1$ to $j = N$) as in equation (1). A good result fulfills:

$$F \approx \sum_{j=1}^{N} Fj \tag{7}$$

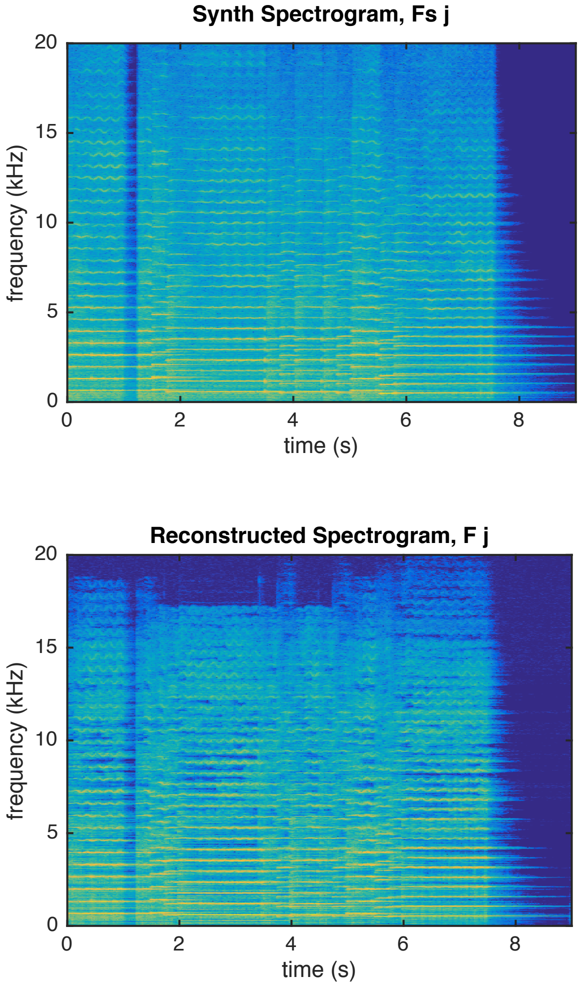
New components (random distributions) can be added to be analyzed in the PLCA reconstruction. In this manner, we can take real sounds that are not present in the score, but they appear in the recording. In addition, you can take residual components that do not belong to any instrument [3].

A more faithful way of adapting these marginal distributions to the mix, is performing a two-step PLCA algorithm. First, we compute only the marginal frequencies $P_j(f|z)$, and in a second step, we iterate to calculate the activations of those frequencies, which are the temporal marginals $P_j(t|z)$. In this case, the second PLCA is used twice, making it a less efficient method.

6. EXPERIMENT AND RESULTS

6.1. Experiment

The experiment is based on an 11-second fragment of "*Entre Dos Aguas*" (1975), a very representative rumba piece of Spanish flamenco by guitarist Paco de Lucía. Fig. 4 shows the score section of this fragment.



4

Figure 3. PLCA adaption of a synth spectrogram (from source j) to the original mix spectrogram *F*, generating a reconstructed spectrogram of the source *j*.

The piece has a fast tempo, flourishes such as *Acciaccaturas* and *Appoggiaturas,* and short notes, which can be very challenging for source separation methods. The fragment consists of a guitar melody accompanied by typical instruments in Spanish folk music, such as the Peruvian Cajon, accompanying guitar and the electric bass. The fragment contains both harmonic and percussive sounds, such as the plucked guitar strings.

The aim of this example is to extract the Spanish guitar melody from the background. The general algorithm described in this work is designed to reconstruct all the instruments from the piece. In this work, we focus on the Spanish guitar, which is the most common instrument in flamenco music. The Spanish guitar (or classical guitar) typically used in Flamenco music has nylon strings (fingerstyle) that result in a sharp attack when plucked. In addition, many factors such as the material and instrument construction affect the spectral content. In the first step, we synthesize the Spanish guitar track from the time-aligned MIDI score with the VST (Virtual Studio Technology) plugin DSK Guitar Nylon.

The synchronism between the MIDI score and the audio is very important for Flamenco music. In a fully automatic version, the algorithm uses the DTW to temporally align the audio with the MIDI. Whenever the DTW presents poor alignment performance, this propagates through to the source separation stage. So, we use manual alignment in this example, which is much more precise than automatic alignment, to allow to focus on the performance of the PLCA in the source separation stage.

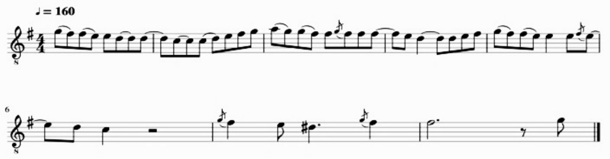


Figure 4. "*Entre Dos Aguas*" Score.

We use an FFT size of 8192 samples with an 88% overlap (7/8 of the FFT). For each PLCA, we do 100 iterations of the EM algorithm. The number of marginals in which we decompose the guitar usually corresponds to the number of different notes played by the instrument along the piece. Each independent sound has a number of marginals that define it. Instruments whose notes have a relatively stable spectrum along the duration of the note require a single marginal (in time and frequency) to define each note. However, Spanish guitar notes don't have a constant spectrum throughout, typically presenting spectro-temporal variations. Thus we define 3 marginal distributions (3 time marginals and 3 frequency marginals) per Spanish guitar note to better capture the frequency variations in time.

6.2. Results and conclusion

Fig. 5 illustrates the result of the Spanish guitar extraction experiment. The top panel shows the spectrogram of the original audio fragment, the middle panel shows the spectrogram of the MIDI synthesis used as prior in PLCA, and the bottom panel shows the spectrogram of the Spanish guitar reconstructed from the PLCA separation. Visual inspection indicates that the melody is extracted and the note attacks correspond with the original voice. However, the sound of the reconstructed guitar is not brilliant enough when we listen to it. Our hypothesis is that the reconstructed guitar fails to capture the characteristic spectro-temporal variation of the Spanish guitar in flamenco music. Plucking the strings of the Spanish guitar results in sharp attacks whose characteristic wide-band spectrum is different from the more harmonic resonant tail end of the notes. The spectrogram of the reconstructed guitar reveals that each note has a constant spectral structure, resulting in attacks that are perceptually closer to the rest of the note.

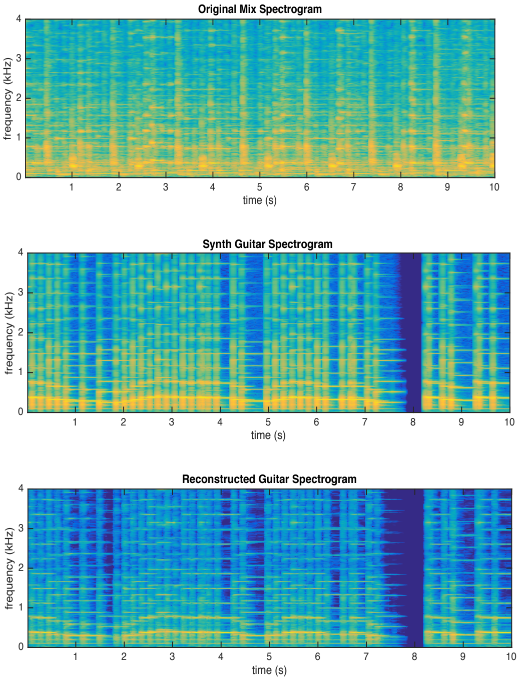


Figure 5. Results of the Guitar Extraction experiment described in 6.1, corresponding to the frequencies under 4 kHz.

The synthesized guitar used as prior in the PLCA greatly influences the spectro-temporal features of the final reconstructed result. If the original guitar synthesis does not contain sharp attacks followed by a more harmonic decay, the PLCA model will not capture spectro-temporal variations in each note. In theory, it is possible to overcome this limitation by introducing additional marginals in the PLCA to capture spectro-temporal variations. However, additional priors would require a means to guarantee that they do not model spectro-temporal variations from the other instruments in the mixture. In this particular example, introducing additional priors in the algorithm degrades the separation because the PLCA model also captures spectro-temporal variations from the instruments in the background.

7. FUTURE WORK

Future work includes the improvement of the modeling of spectro-temporal variations to further improve the results of the source separation for the challenging sharp attacks of the Spanish Guitar. In addition, it is useful to implement a method of automatic transcription of scores, since many works of flamenco are not written, as well as to compare results of using other methods of source separation, as methods of blind separation that do not require the use of a score.

8. ACKNOWLEDMENTS

This work has been funded by Ministerio de Economía y Competitividad of the Spanish Government under Project No. TIN2016-75866-C3-2-R. This work has been done at Universidad de Málaga, Campus de Excelencia Internacional Andalucía Tech.

9. REFERENCES

[1] L. Benaroya, F. Bimbot, L. McDonagh, R. Gribonval (2003). "Non negative sparse representation for Wiener based source separation with a single sensor". In *IEEE Int. Conf. Audio Speech Signal Process, (*pp. 613-616).

[2] D. FitzGerald (2004). "Automatic drum transcription and source separation". Dublin Inst. Technol.

[3] Joachim Ganseman, Paul Scheunders, Gautham J. Mysore, Jonathan S. Abel. (2010). "Source separation by score synthesis". *In Proceedings of the International Computer Music Conference (ICMC)*, New York, USA, (pp. 462–465).

[4] M. Heln, T. Virtanen (2005). "Separation of drums from po-lyphonic music using nonnegative matrix factorization and support vector machine". In *Proc. Eur. Signal Process. Conf.*

[5] M. Shashanka, B. Raj, P. Smaragdis (2008). "Probabilistic latent variable models as non-negative factorizations." In *Computational Intelligence and Neuroscience Journal, special issue on Advances in Non-negative Matrix and Tensor Factorization*.

[6] Ning Hu, Roger B. Dannenberg, and George Tzanetakis (2003). "Polyphonic audio matching and alignment for music retrieval". In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA.

[7] Orio, N., Schwarz (2001). "Alignment of Monophonic and Polyphonic Music to a Score". In *Proc. 2001 ICMC,* (pp. 155-158).

[8] Paris Smaragdis, Bhiksha Raj, madhusudana Shashanka (2006). "Supervised and Semi-Supervised Separation of Sounds from Single-Channel Mixtures". In *International Conference on Independent Component Analysis and Signal Separation*.

[9] Romain Hennequin, Bertrand David, Roland Badeau (2011). "Score informed audio source separation using a parametric model of non-negative spectrogram". In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, (pp. 45–48).

[10] Sebastian Ewert, Meinard Müller (2012). "Using score-informed constraints for NMF-based source separation". In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan.

[11] T. Virtanen (2003). "Sound source separation using sparse coding with temporal continuity objective". In *Proc. Int. Comput. Music Conf.*, (pp. 231-234).

[12] T. Virtanen (2004). "Separation of sound sources by convolutive sparse coding". In *Proc. ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Process*.

[13] Y. Guo and M. Zhu (2011). "Audio source separation by basis function adaptation". In *Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague. IEEE, (pp. 2192-2195).

[14] Zhiyao Duan, Bryan Pardo (2011). "Soundprism: An online system for score-informed source separation of music audio". In *IEEE Journal of Selected Topics in Signal Processing,* (pp. 1205–1215).

SCORE-INFORMED SYLLABLE SEGMENTATION FOR JINGJU A CAPPELLA SINGING VOICE WITH MEL-FREQUENCY INTENSITY PROFILES

Rong Gong
Music Technology Group,
Universitat Pompeu Fabra,
Barcelona, Spain
rong.gong@upf.edu

Nicolas Obin
IRCAM, CNRS,
UPMC-Sorbonne Universités,
Paris, France
nicolas.obin@ircam.fr

Georgi Dzhambazov, Xavier Serra
Music Technology Group,
Universitat Pompeu Fabra,
Barcelona, Spain
georgi.dzhambazov@upf.edu
xavier.serra@upf.edu

ABSTRACT

This paper introduces a new unsupervised and score-informed method for the segmentation of singing voice into syllables. The main idea of the proposed method is to detect the syllable onset on a probability density function by incorporating *a priori* syllable duration derived from the score. Firstly, intensity profiles are used to exploit the characteristics of singing voice depending on the Mel-frequency regions. Then, the syllable onset probability density function is obtained by selecting candidates over the intensity profiles and weighted for the purpose of emphasizing the onset regions. Finally, the syllable duration distribution shaped by the score is incorporated into Viterbi decoding to determine the optimal sequence of onset time positions. The proposed method outperforms conventional methods for the segmentation of syllable on a jingju (also known as Peking or Beijing opera) *a cappella* dataset. An analysis is conducted on precision errors to provide direction for future improvement.

1. INTRODUCTION

1.1 Context and motivations

Indication from both psychoacoustic and psycholinguistical research Massaro (1974); Segui et al. (1990); Greenberg (1996) suggests that the syllable is a basic perceptual unit for speech processing in humans. The syllable was recommended as a basic unit of automatic speech recognition as early as 1975 Mermelstein (1975). The syllabic level offers several potential benefits; for one, contrary to the phoneme system which is specific to a language, the syllable is universally defined in terms of acoustic sonority [1] : a syllable segment is fully determined by a maximum of sonority (the vowel nucleus) surrounded by local minima of sonority. Additionally, the syllable is the basic unit of the prosody analysis of speech or singing voice.

In contrast to speech syllables, the duration of singing voice syllables varies enormously and their vowel nucleus may consists of numerous local sonority maxima due to the various ornaments, typically the vibrato - amplitude and frequency modulation, which poses new challenge for the segmentation task. A musical score contains a wide range of prior information, such as the pitch, the onset time and the duration of the note and the syllable, which can be used to guide the segmentation process.

1.2 Related work

Most of existing speech syllable segmentation methods can be divided into two categories: unsupervised Mermelstein (1975); Wang & Narayanan (2007); Obin et al. (2013) and supervised Howitt (2000); J. Makashay et al. (2000). In the Mermelstein method Mermelstein (1975), the syllable onset are detected by recursively searching on the convex hull of the loudness function. Wang & Narayanan (2007) have explored the Mel-frequency spectral representations for syllable segmentation. Most recently, the Syll-O-Matic system Obin et al. (2013) exploited the fusion of Mel-frequency intensity profiles and voicing profiles which gives the best segmentation result for the methods of the first category. Supervised methods Howitt (2000); J. Makashay et al. (2000) adopted from Automatic Speech Recognition need the support of a language model and an acoustic model. The latter is learned from a set of audio recordings and their corresponding transcripts, which takes a considerable amount of time to adapt this method from one language to another.

The syllable segmentation of singing voice is still a research gap which needs to be filled. The related subjects are singing voice phonetic segmentation Lin & Jang (2007), lyrics-to-audio alignment Fujihara & Goto (2012); Dzhambazov et al. (2016), and score-to-audio alignment of singing voice Gong et al. (2015). The approaches adopted in these works are mostly supervised, so the problems of the language specificity and the need for a large amount of training data remain.

Various applications such as score-informed source separation Ewert et al. (2014); Miron et al. (2015), tonic identification Sentürk et al. (2013) and score-to-audio alignment Cont (2010) have been proposed in recent years which exploit the availability of a musical score. Dzhambazov et al. (2016) shows that modeling of duration improves the phrase-level lyrics-to-audio alignment accuracy significantly.

This paper introduces a new unsupervised and score-informed method for the segmentation of singing phrase into syllables. We present the definitions of speech syllable and jingju singing voice syllable, and disclose the issues existing in syllable segmentation in section 2. The

[1] the relative loudness of a speech sound.

ISBN: 978-84-697-2303-6

approach is explained in section 3. The evaluation and the error analysis are conducted on a jingju *a cappella* singing voice dataset in section 4.

2. WHAT IS A SYLLABLE?

2.1 Definition

The task of automatically detecting the speech syllable is based on the assumption that a syllable is typically vowel centric and neighboring vowels are always separated by consonants Howitt (2000). A precise characterization of the syllable structure can be made in terms of sonorityAssociation (1999), which hypothesizes that syllables contain peaks of sonority that constitute their nuclei and may be surrounded by less sonorous sounds Goldsmith et al. (2011). According to the Sonority Sequencing Principle Dressler (1992), vowels and consonant sounds span a sonority continuum with vowel nuclei being the most sonorous and obstruents being the least, with glides, liquids, and nasals in the middle.

Mandarin is a tonal language and there are in general 4 lexical tones and 1 neutral tone in it. Every character of spoken Mandarin language is pronounced as monosyllable Lin et al. (1993). The jingju singing is the most precisely articulated rendition of the spoken Mandarin language. Although certain special pronunciations in jingju theatrical language differ from their normal Mandarin pronunciations, due to firstly the adoption of certain regional dialects, and secondly the ease or variety in pronunciation and projection of sound, the mono-syllabic pronouncing structure of the standard Mandarin doesn't change Wichmann (1991).

A syllable of jingju singing is composed of three distinct parts in most of the cases: the "head" (tou), the "belly" (fu) and the "tail" (wei). The head consists of the initial consonant or semi-vowel, and the medial vowel if the syllable includes one, which itself is normally not prolonged in its pronunciation except for the one with a medial vowel. The belly follows the head and consists of the central vowel. It is prolonged throughout the major portion of the melodic-phrase for a syllable. The belly is the most sonorous part of a jingju singing syllable and can be analogous to the nuclei of a speech syllable. The tail is composed of the terminal vowel or consonant Wichmann (1991).

The speech syllable only contains one prominent sonority maximum due to its short duration (average < 250 ms and standard deviation < 50 ms for Mandarin Wang (1994)). In contrast, a singing voice syllable may consists of numerous local sonority maxima, of which the reason is either intentional vocal dynamic control for the needs of conveying a better musical expression or unintentional vocal intensity variation as a by-product of the F0 change Titze & Sundberg (1992).

2.2 Issues in syllable segmentation

The issues of speech syllable segmentation has been summarized in Obin et al. (2013). Jingju singing voice brings up two new issues. Firstly, the syllable duration of jingju singing voice varies enormously. According to the statistics of our dataset, the syllable durations range from 70 ms to 21.7 s and its standard deviation is 1.74 s, which makes it impossible to model the durations with one single distribution as it has been done for speech Obin et al. (2013). Secondly, as mentioned in section 2.1, the syllable's central vowel may consists of numerous local sonority maxima, which introduces noisy information for the syllable segmentation.

A priori syllable duration information is often easy to obtain from the score and this is an advantage which can be exploited. The repertoire of jingju includes around 1400 plays Wichmann (1991), among which are still performed and used in teaching are mostly well transcribed into sheet music. Constructing the syllable duration distribution from the score and using it to guide the segmentation process is a feasible way of solving the two new issues mentioned above.

3. APPROACH

The objective of this study is automatically segmenting singing phrases into syllables by incorporating syllable duration information derived from the score into syllable onset detection. Firstly, Mel-frequency intensity profiles are measured over various frequency regions. An observation probability function of syllable onsets is obtained by selecting candidates over the intensity profiles and weighted for the purpose of augmenting its value in the onset regions. Secondly, the *a priori* duration distribution derived by the score is incorporated into the Viterbi decoding to determine the optimal sequence of syllabic onset time positions (Fig.1). The conventional unsupervised speech syllable segmentation method is based on the detection of syllable onset and landmark Obin et al. (2013). However, we focus the issue only on onset detection because the definition of syllable landmark Howitt (2000) doesn't apply to jingju singing voice due to the numerous local sonority maxima within the central vowel.

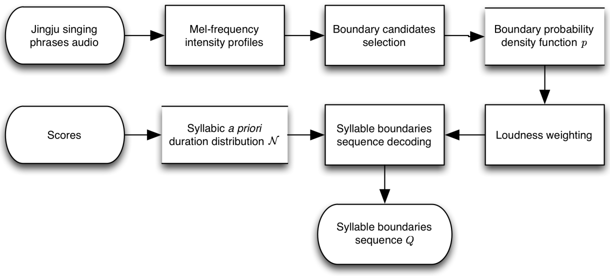

Jingju singing phrases audio
Mel-frequency intensity profiles
Boundary candidates selection
Boundary probability density function p
Scores
Syllabic *a priori* duration distribution $\mathcal{N}$
Syllable boundaries sequence decoding
Loudness weighting
Syllable boundaries sequence Q

Figure 1: Approach diagram.

3.1 Mel-frequency intensity profiles

A time-frequency representation is used to measure the intensity contained into various frequency regions. For each frequency region, the specific loudness is measured as:

$$\mathrm{L}_t^{(k)} = \sum_{n=1}^{N^{(k)}} {|A(t,n)|^2}^{\,0.23} \qquad (1)$$

where k denotes the k-th frequency region, $A(t,n)$ the amplitude of the n-th frequency bin at time t in the considered frequency region, and $n = 1$ the start value of the summation index in the k-th frequency region. The specific loudness is related to the sound intensity - the square of the amplitude $A(t,n)$ through a power law with an exponent 0.23 Zwicker & Fastl (2013). In this study, the specific loudness is measured over 40 Mel-frequency bands, with unitary integrated energy in order to enhance the information contained in low-frequency regions relatively to high-frequency regions. The frequency bands are equally spaced on the mel scale Slaney (1998), which approximates the human auditory system's response more closely than the linearly-spaced frequency bands. Then, the specific loudness $\mathrm{L}_t^{(k)}$ is normalized into a probability density function ${\mathrm{L}_t^{(k)}}_{\text{norm}}$ so that each intensity profile will be further equally processed (Fig.2-b).

3.2 Onset candidates selection

A syllable has a great probability of starting with a consonant. Stop consonants consist of an interval of complete closure. Because of this, all stops have a period of silence. Affricates consonants have frication portion preceded by stop-like 'silent' portion. Liquids consonants are normally voiced, but have less energy than vowels Johnson (2011). Accordingly, consonants, apart from fricatives and nasals, contain a complete silence or less energy (intensity) than vowels. Additionally, a syllable is usually preceded by some silence or breath frames which also have low intensities in certain frequency regions. These characteristics incite us to conduct the syllabic onset detection on ${\mathrm{L}_t^{(k)}}_{\text{norm}}$ by a local maxima-minima detection method Obin et al. (2013), which gives a local minima onset candidate sequence $\mathrm{Onset}^{(k)}$ for each Mel-frequency band k.

The local maxima-minima detection method consists of two steps: in the first step, we conduct a coarse search to find all the maxima and minima positions; in the second step, the positions are selected such that the maxima are required to exceed both neighboring minima by at least a heuristic height threshold (0.01 relative amplitue) and to be separated by at least an heuristic offset threshold (0.025s); otherwise, the maxima together with their neighboring minima are considered as insignificant and suppressed.

This process forms a (K × T) matrix of onset time-frequency position candidates (Fig.2-c). K and T denote respectively the numbers of the Mel-frequency bands and the time frames. Then, it is summed up into a (1 × T) probability density function p (Fig.2-d) because the more frequent is observed a time position of a candidate over frequency bands, the more likely is the presence of an onset. However, the exact time position of an onset may differ from one frequency region to the other due to the asynchronism of the information contained in the frequency regions. Thus, a moving average window MA (typically, a 20 ms. window) is employed.

$$p = \mathrm{MA}(\sum_{k=1}^{K} \mathrm{Onset}^{(k)}) \qquad (2)$$

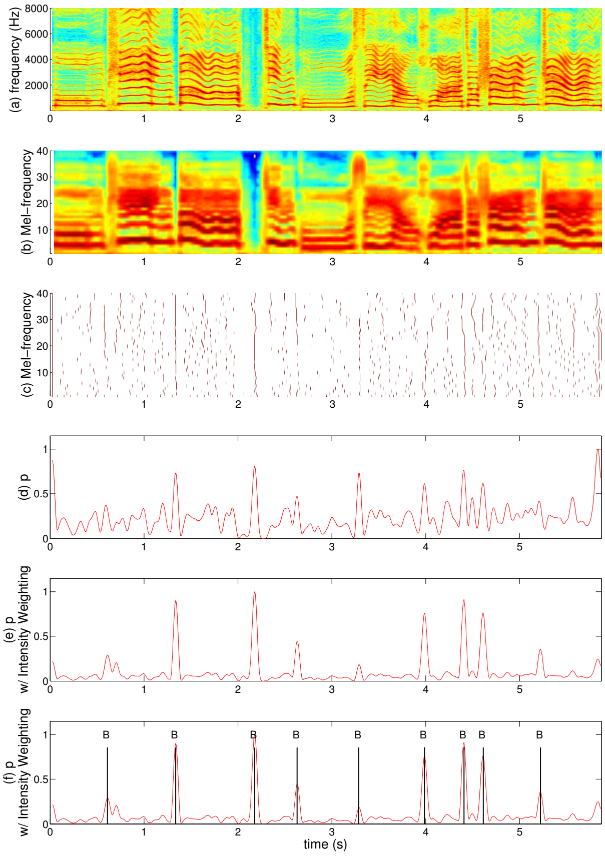


Figure 2: Spectrogram (a), Mel-frequency intensity profiles (b), (K × T) matrices of onset time/frequency position candidates (c), onset probability density function p (d) and its loudness weighted version (e), determined sequence of syllable onsets (f) for the singing phrase: "Meng ting de jin gu xiang hua jiao sheng zhen."

3.3 Loudness weighting

Certain prominent peak positions can be identified as the syllable onsets on the graph of the probability density function p (Fig. 2-d). However, numerous less prominent peaks can also be found, which do not correspond to the real syllable onsets. This noisy information (less prominent peaks) will eventually degrade the performance of the onset sequence decoding. By observing the graph (Fig.2-d), we clearly see that most of these noisy peaks appear in the vowel regions which usually show a high intensity Dressler (1992). To reduce these noisy peaks, we scale down the high-intensity regions of p by multiplying it by a weighting coefficient.

Inspired by the loudness gating method used in EBU (2016), we employ an absolute gating threshold θ_a, a relative gating threshold θ_r and a sound pressure level storing block SPL_i to detect the high-intensity signal frames.

The intensity of the input singing voice signal is measured frame by frame by the sound pressure level of its RMS amplitude $SPL_{\text{RMS}} = 20\log_{10}(RMS)$. The current frame is detected as high-intensity if its SPL_{RMS} meets both of the following conditions:

$$SPL_{\text{RMS}} \geqslant \theta_a \tag{3}$$

$$SPL_{\text{RMS}} \geqslant \theta_r + \overline{SPL_i} \tag{4}$$

where $\overline{SPL_i}$ is the mean value of the integrated preceding stored SPL_{RMS}, θ_a, θ_r are heuristically selected as -35 dB and -10 dB. Once a frame is detected as high-intensity, its SPL_{RMS} is added to the storing block SPL_i.

A continuous sequence of high-intensity frames is detected as the high-intensity region if it is followed by a continuous sequence of low-intensity frames. The length of the latter should be larger than a threshold θ_l which will be optimized by the grid search method. Finally, the p value in the high-intensity regions is multiplied by a weighting coefficient w_h which will also be optimized later. (Fig.2-e).

3.4 A priori duration distribution

The *a priori* duration distribution $\mathcal{N}(x; \mu_l, \sigma_l^2)$ is modeled by a Gaussian function whose mean μ_l equals to l-th syllable duration of the score and whose standard deviation σ_l is proportional to μ_l: $\sigma_l = \gamma\mu_l$ (Fig.3). The proportionality constant γ will be optimized by the grid search method.

$$\mathcal{N}(x; \mu_l, \sigma_l^2) = \frac{1}{\sqrt{2\pi}\sigma_l} \exp\left(-\frac{(x-\mu_l)^2}{2\sigma_l^2}\right). \tag{5}$$

The relative duration of each note is measured on the quarter note length, so an eighth note has a duration of 0.5. We only keep the relative duration and discard the tempo information of the score. By normalizing the summation of the notes' relative durations to unity, then multiplying it by the duration of the incoming audio recording, we obtain the absolute score duration of the entire singing phrase which is equal to the latter. The note's absolute duration along with its subsequent silence or the summation of the notes' absolute durations (e.g. syllable *gu* in Fig.3) corresponding to l-th syllable is assigned to μ_l. The duration distribution (Eq.5) will be incorporated into Viterbi algorithm as the state transition probability, which holds the highest expectation on its mean value - the syllable duration of the score.

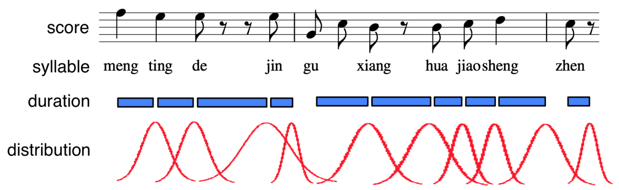


Figure 3: *A priori* relative duration distributions (bottom) of the syllables in the singing phrase: "Meng ting de jin gu xiang hua jiao sheng zhen."

3.5 Decoding of the syllable onsets sequence

A sequence of *a priori* absolute duration $M = \mu_1\mu_2\cdots\mu_L$ is deduced from the score and the length of the incoming audio (section 3.4). To decode the syllable boundaries, we construct an hidden Markov model characterized by the following:

1. The state space is a set of N candidate onset positions $S_1, S_2, \cdots, S_N$ determined by picking the local maxima positions from the probability function p.
2. The state transition probability at decoding time l is defined by *a priori* duration distribution $\mathcal{N}(d_{ij}; \mu_l, \sigma_l^2)$, where d_{ij} is the time distance between states S_i and S_j $(j > i)$. The overall decoding time is equal to the total syllable number L written in the score.
3. The observation probability for the state S_j is represented by its corresponding value in the onset detection function p, which is denoted as p_j.

As we assume the onset of the current syllable is also the offset of the previous syllable, the problem is translated into finding the best offset position state sequence $Q = q_1q_2\cdots q_L$, for the given *a priori* duration sequence M, where q_i denotes the offset of the ith decoding syllable or the onset of the $i+1$th decoding syllable. q_0 and q_L are fixed as S_1 and S_N as we expect that the onset of the first syllable is located in the beginning of the incoming audio and the offset of the last syllable is located in the ending of the audio. One can fulfill this assumption by truncating the silences at both ends of the incoming audio. According to the logarithmic form Viterbi algorithm Rabiner (1989), we define

$$\delta_l(i) = \max_{q_1, q_2, \cdots, q_l} \log P[q_1q_2\cdots q_l,\ \mu_1\mu_2\cdots\mu_l]$$

the initially step

$$\delta_1(i) = \log(\mathcal{N}(d_{1i}; \mu_1, \sigma_1^2)) + \log(p_i)$$
$$\psi_1(i) = S_1$$

the recursion step

$$\delta_l(j) = \max_{1\leqslant i<j}[\delta_{l-1}(i) + \log(\mathcal{N}(d_{ij}; \mu_l, \sigma_l^2))] + \log(p_j)$$
$$\psi_l(j) = \arg\max_{1\leqslant i<j}[\delta_{l-1}(i) + \log(\mathcal{N}(d_{ij}; \mu_l, \sigma_l^2))]$$

and termination step

$$\log P^* = \max_{1\leqslant i<N}[\delta_{L-1}(i) + \log(\mathcal{N}(d_{iN}; \mu_L, \sigma_L^2))]$$
$$q_L^* = \arg\max_{1\leqslant i<N}[\delta_{L-1}(i) + \log(\mathcal{N}(d_{iN}; \mu_L, \sigma_L^2))]$$

Finally, the best offset position state sequence Q is obtained by the backtracking step (Fig.2-f).

4. EVALUATION

4.1 Dataset

The *a cappella* singing dataset [2] used for this study comes from MTG and C4DM Black et al. (2014) and focuses

[2] http://doi.org/10.5281/zenodo.345490

on two most important jingju role-types Repetto & Serra (2014): *dan* (female) and *laosheng* (old man). It contains 39 interpretations of 31 unique arias sung by 11 jingju singers. The syllable onset ground truth is manually annotated in Praat Boersma (2001), which represents 298 phrases and 2672 syllables (including padding written - characters Wichmann (1991)). The average syllable duration is 1.1s and the standard deviation is 1.74s. The syllable duration dataset is manually transcribed from sheet music.

The whole dataset is randomly split into 2 parts with the constraint that each part is selected without role-type bias and contains almost an equal number of onsets. One of them is reserved as the development set for the purpose of parameter optimization. Another part is used as the test set to evaluate the syllable segmentation algorithms.

4.2 Evaluation metrics

The objective of the syllable segmentation for singing phrases is to determine the time positions of syllable boundaries. The evaluation consisted in the comparison of the determined syllable onsets and offsets to the reference one. We use the same metric for the speech syllable segmentation evaluation: recall, precision and F-measure Obin et al. (2013). The definition of a correct segmented syllable is borrowed from the note transcription evaluation Molina et al. (2014): for the syllable onset, we choose a evaluation tolerance $\pm\tau$ ms. For the offset, which is also the onset of the subsequent syllable, $\pm 20\%$ of the reference syllable's duration or $\pm\tau$ ms, whichever is larger, is chosen as the tolerance. If both the onset and the offset of a syllable lie within the tolerance of their reference counterparts, we say it's correctly segmented. As there is no standard tolerance previously defined for the evaluation of singing voice syllable onset detection, and the tolerance for the evaluation of speech syllable onset detection is too strict because the average duration of speech syllable (200 ms) is much shorter than that of singing voice syllable (1.1 s), we decide to report the evaluation results for multiple tolerances, $\tau = [0.05, 0.1, 0.15, 0.2, 0.25, 0.3]$ (second).

4.3 Parameters optimization

The parameters which need to be optimized are: the length threshold θ_l of low-intensity regions, the weighting coefficient w_h for p in high-intensity regions in section 3.3; the proportionality constant γ in section 3.4. The syllable segmentation accuracy can be reported by sweeping these parameters on the development set. Table 1 lists the search bounds and the optimal results.

Table 1: Search bounds, optimal results (OR) of the optimization process for each parameter.

Parameters	Search bounds	OR
θ_l (s)	$[0.01, 0.1]$ with step 0.01	0.02
w_h	$[0.1, 1]$ with step 0.1	0.2
γ	$[0.05, 1]$ with step 0.05	0.35

5. RESULTS AND DISCUSSION

5.1 Syll-O-Matic syllable segmentation

The evaluation includes the speech syllable segmentation method Syll-O-Matic Obin et al. (2013) for a comparison with the unsupervised method. This method performs the same Mel-frequency intensity profiles and onset candidate selection steps introduced in this paper. It detects both the speech syllable onsets and the vowel landmarks. We will not report its landmark detection performance because the definition of syllable landmark - the only and most sonorous peak with the central vowel, doesn't apply to most of jingju singing syllables due to the existence of numerous local sonority maxima within the central vowel.

Our proposed method can be seen as an adaption of the original Syll-O-Matic method to the singing voice, which introduces the loudness weighting to attenuate the noisy peaks in the onset probability density function p, and *a priori* syllable duration distribution to take account into the duration information provided by the score, whereas only a fixed mean (1.1s, the average syllable duration of our dataset) normal distribution has been used in the Viterbi decoding process of the original Syll-O-Matic method.

The Syll-O-Matic method performs bad on our dataset (Fig.4) and causes a low F-measure. There are at least three reasons for this bad performance. First, its Viterbi decoding algorithm doesn't restrict the overall decoding time, so any peak position in p is able to be decoded as a syllable onset if it happens to have a high duration probability. Second, the numerous sonorous peaks in p act as the noisy information, which introduces many insertions. Third, the duration distribution used in Syll-O-Matic is mean-fixed, which doesn't conform to the fact of the variable syllable duration of the jingju singing voice.

5.2 HMM-based lyrics-to-audio alignment

The evaluation also includes a HMM-based lyrics-to-audio alignment method Dzhambazov et al. (2016) for a comparison with the supervised method. The HMM-based system extends Viterbi decoding to handle the duration of states. For each of 40 Mandarin phonemes and diphthongs, a one-state HMM is trained from a 67 minutes corpus of *a cappella* female jingju singing voice. This corpus is different from the one mentioned in section 4.1 in terms of the singer and the repertoire. For each state a 40-mixtures of Gaussian distribution are fitted on the MFCCs feature vector. The HMM-based system outputs the decoded syllable onset positions.

5.3 Proposed method

Even without the loudness weighting step, the proposed method (Score-informed) outperforms all the compared methods. Additionally, the loudness weighting (Score-informed + Loudness weighting) successfully improves the segmentation performance due to the reduction of the noisy information in the high-intensity region of the probability density function. Compared to supervised methods (e.g. HMM-based), the results are encouraging for the use of

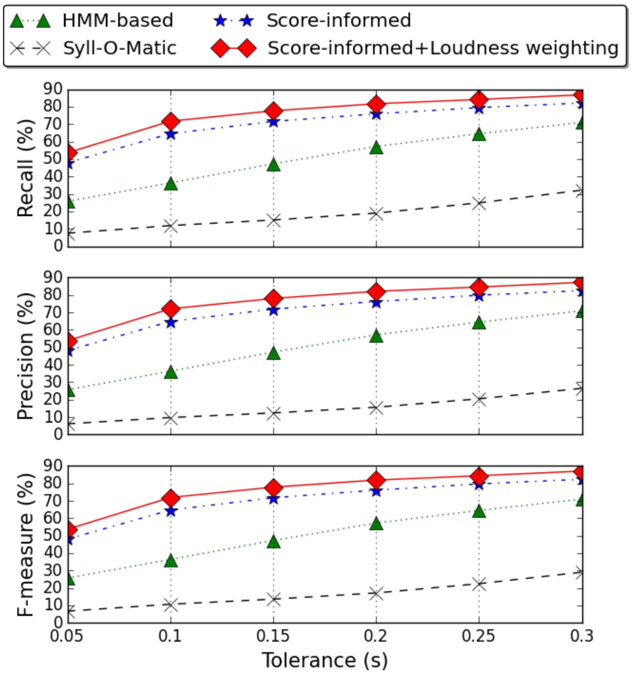


Figure 4: Recall, precision and F-measure results of the syllable segmentation evaluation. The three metrics do not look so different because the number of the segmented syllable and the number of the ground truth syllable are almost the same.

unsupervised score-informed method for singing voice syllable segmentation, which avoids the problems of the language specificity and the need for a large amount of training data.

5.4 Error analysis

We conduct error analysis to make clear the causes of segmentation errors of our proposed method, and also to provide direction for future improvement. First, only the errors occurred in the segmented syllables (precision errors) will be analyzed because the number of the segmented syllable (1329) and the number of the ground truth syllable (1334) are almost equal for the result of the proposed method, which means almost all the syllables in the ground truth are segmented. Second, only the errors out of 0.3s tolerance will be analyzed because the causes of these errors are straightforward to be identified from observing the segmentation plots. 169 syllables are mistakenly segmented out of 1329 evaluated syllables (Table 2).

Table 2: Performance of the proposed method with 0.3s tolerance.

Method	Recall(%)	Precision(%)	F-measure(%)
Score-informed+ Loudness weighting	86.83	87.28	87.05

Four types of error have been identified (Table 3) by observing the plots of detected syllable onsets compared to ground truth onsets:

- Redundant intensity minima: errors caused by redundant intensity minima (redundant peaks) in the onset probability density function p. Silence or large intensity change within the syllable are the main causes of this error type.
- Missed intensity minima: errors caused by missed intensity minima (missed peaks) in p. Long silence followed by the syllable is the main cause of this error type, which usually happens in *laosheng* (old man) singing.
- Ambiguous syllable transitions: errors caused by ambiguous syllable transitions, such as transitions from vowel to vowel or to semi-vowel, from semi-vowel to semi-vowel. This cause has also been reported in the unsupervised speech syllable segmentation research Obin et al. (2013).
- Score and singing incoherent: errors caused by large contrast between syllable duration in score and that in real practice.

Table 3: Error analysis for the result of the proposed method with 0.3s tolerance.

Types of error	Num. errors (frequency %)
Redundant intensity minima	92 (54.3)
Missed intensity minima	34 (20.3)
Ambiguous syllable transitions	32 (18.8)
Score and singing incoherent	11 (6.6)
Sum	169 (100)

The reason for the first three types of error is that our proposed method only uses intensity-related feature and technique (Mel-frequency intensity profiles and loudness weighting) which are not knowledgeable in the phonetic context of the signal frames. By applying phonetic features to shape the peaks of the onset probability density function in the future, for example, comparing the phonetic content before and after the silence, we may reduce these types of error. For the last type of error - Score and singing incoherent, the effort should be put in improving the onset decoding method. Using different duration distribution function, such as gamma distribution, and variable decoding time can be the possible way to tackle this type of error.

6. CONCLUSION

In this paper, we present the definition of jingju singing voice syllable and disclose the new issues arose by this singing form. A new method is then introduced for the segmentation of singing voice into syllables. The main idea of the proposed method is to detect the syllable onset on a syllable onset probability density function by incorporating the syllable duration information of the score into the decoding process. The main contribution of this work is twofold: First, the loudness weighting is applied on the high-intensity regions of the onset probability density function, which reduced the noisy sonorous peaks and augmented the segmentation accuracy. Second, the syllable duration distribution is incorporated into the decoding process of the optimal syllable onset sequence to make use of the *a priori* information of the score. The proposed method outperforms conventional methods for the syllable segmentation of singing voice phrases, and provides a promising paradigm for the segmentation of singing voice into syllables.

7. ACKNOWLEDGEMENTS

This work is partly supported by the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grand agreement 267583).

8. REFERENCES

Association, I. P. (1999). *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge University Press.

Black, D. A. A., Li, M., & Tian, M. (2014). Automatic Identification of Emotional Cues in Chinese Opera Singing. In *ICMPC-2014*, Seoul, South Korea.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Cont, A. (2010). A coupled duration-focused architecture for realtime music to score alignment. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, *32*(6), 974–987.

Dressler, W. (1992). *Phonologica 1988*. Cambridge University Press.

Dzhambazov, G., Yang, Y., Repetto, R. C., & Serra, X. (2016). Automatic alignment of long syllables in a cappella beijing opera. In *FMA-2016*, Dublin, Ireland.

EBU (2016). 'EBU Mode' metering to supplement Loudness normalisation. Recommendation Tech 3341-2016, Geneva. Version 3.0.

Ewert, S., Pardo, B., Muller, M., & Plumbley, M. (2014). Score-Informed Source Separation for Musical Audio Recordings: An overview. *IEEE Signal Processing Magazine*, *31*(3), 116–124.

Fujihara, H. & Goto, M. (2012). Lyrics-to-audio alignment and its application. *Dagstuhl Follow-Ups*, *3*.

Goldsmith, J. A., Riggle, J., & Yu, A. C. L. (2011). *The Handbook of Phonological Theory*. John Wiley & Sons.

Gong, R., Cuvillier, P., Obin, N., & Cont, A. (2015). Real-Time Audio-to-Score Alignment of Singing Voice Based on Melody and Lyric Information. In *Inter Speech-2015*, Dresden, Germany.

Greenberg, S. (1996). *Understanding Speech Understanding: Towards A Unified Theory Of Speech Perception*.

Howitt, A. W. (2000). *Automatic Syllable Detection for Vowel Landmarks*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.

J. Makashay, M., W. Wightman, C., K. Syrdal, A., & Conkie, A. (2000). Perceptual Evaluation of Automatic Segmentation in Text-to-Speech Synthesis. In *ICSLP 2000*, Beijing.

Johnson, K. (2011). *Acoustic and Auditory Phonetics*. John Wiley & Sons.

Lin, C.-H., Lee, L.-s., & Ting, P.-Y. (1993). A new framework for recognition of Mandarin syllables with tones using sub-syllabic units. In *ICASSP-1993*, volume 2.

Lin, C.-Y. & Jang, J.-S. (2007). Automatic Phonetic Segmentation by Score Predictive Model for the Corpora of Mandarin Singing Voices. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*(7).

Massaro, D. W. (1974). Perceptual units in speech recognition. *Journal of Experimental Psychology*, 199–208.

Mermelstein, P. (1975). Automatic segmentation of speech into syllabic units. *The Journal of the Acoustical Society of America*, *58*(4), 880–883.

Miron, M., Carabias-Orti, J. J., & Janer, J. (2015). Improving Score-Informed Source Separation for Classical Music through Note Refinement. In *ISMIR-2015*, Malaga.

Molina, E., Barbancho, A. M., Tardón, L. J., & Barbancho, I. (2014). Evaluation Framework for Automatic Singing Transcription. In *ISMIR-2014*, Taipei, Taiwan.

Obin, N., Lamare, F., & Roebel, A. (2013). Syll-O-Matic: An adaptive time-frequency representation for the automatic segmentation of speech into syllables. In *ICASSP-2013*.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286.

Repetto, R. C. & Serra, X. (2014). Creating a Corpus of Jingju (Beijing Opera) Music and Possibilities for Melodic Analysis. In *ISMIR-2014*, Taipei, Taiwan.

Segui, J., Dupoux, E., & Mehler, J. (1990). *Cognitive Models of Speech Processing*. Cambridge, MA, USA: MIT Press.

Sentürk, S., Gulati, S., & Serra, X. (2013). Score informed tonic identification for makam music of turkey. In *ISMIR-2013*.

Slaney, M. (1998). Auditory toolbox. Technical Report 1998-010, Interval Research Corporation.

Titze, I. R. & Sundberg, J. (1992). Vocal intensity in speakers and singers. *The Journal of the Acoustical Society of America*, *91*(5), 2936–2946.

Wang, D. & Narayanan, S. (2007). Robust Speech Rate Estimation for Spontaneous Speech. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*(8), 2190–2201.

Wang, J. (1994). Syllable duration in Mandarin. In *the Fifth International Conference on Speech Science and Technology*.

Wichmann, E. (1991). *Listening to Theatre: The Aural Dimension of Beijing Opera*. University of Hawaii Press.

Zwicker, E. & Fastl, H. (2013). *Psychoacoustics: Facts and models*, volume 22. Springer Science & Business Media.

WHAT IF FIDDLING SOLVES YOUR PROBLEMS?

Enric Guaus, Oriol Saña
Music Research and Creation Group,
Escola Superior de Música de Catalunya
[enric.guaus,oriol.sana]@esmuc.cat

Laura Ramos
Conservatorio Superior de Música de Castellón
laura_rs1@hotmail.com

ABSTRACT

It is well known that learning how to play a musical instrument is a difficult task. During this long process, some decisions may help the students improve their performing skills. In the case of violin students, and in the context of the Escola Superior de Música de Catalunya (ESMUC), some teachers detected that students taking some lessons in the folk tradition style *play better* with respect to those students who don't take these lessons, when playing specific repertoire from the classical tradition. But, is it possible to quantify this improvement? What do teachers exactly mean when they say students play *better*? This study shows a methodology to quantitatively differentiate between these two groups of students, and discusses which musical aspects define this improvement.

1. INTRODUCTION

The main goal of this research is to empirically verify that violin students from the classical tradition achieve a significant improvement in the performance of their common repertoire when applying learning methods and techniques used in the folk tradition (i.e. Fiddle).

For that, we need to include several areas of knowledge ranging from musicological (what does *performance improvement* mean?), pedagogical (how can we design and evaluate a learning process?) to engineering (which techniques for Motion Capture (MOCAP) should we use?) and statistical analysis (which descriptors from Music Information Retrieval (MIR) should we extract?).

This work is divided in three parts. Section 2 debates about the context of this study and discusses about the facets of music to be analyzed. Section 3 presents a detailed description of the recording sessions and technical setup for data acquisition, and Section 4 shows the results of the preliminary analysis and a verification of these results in a new bench of measurements. At the end, some conclusions about pedagogical implications derived from this study are presented in addition to some remarks about research reproducibility.

2. PERFORMANCE SKILLS

First, we discuss about what we understand as *an improvement in the performance skills*. For that, we need to tackle the concept of *quality* in the performance. In general, a musician is able to adapt the performance of a given score in order to achieve certain musical and emotional effects, that is, provide an *expressive musical performance*. There exists a huge literature for the analysis of expressive musical performances. Focusing on the approaches using technologies, Widmer & Goebl (2004) provides a good overview on this topic. Under our point of view, one of the most relevant contributions is the Performance Worm for the analysis of performances by Dixon et al. (2002). It shows the evolution of tempo and perceived loudness information in a 2D space in real time, with a decreasing brightness according to a negative exponential function to show past information. Saunders et al. (2004) analyzed the playing styles from different pianists using (beat-level) tempo and (beat-level) loudness information. In addition to that, different systems have been developed to allow machines create expressive music, which are summarized by Kirke & Reck Miranda (2009). In summary, most of the studies related to expressive performances are based on loudness and rhythmic properties of music. According to our main goal which is to present evidences in differences of performances for violin students from fiddle and classical traditions, in the context of a music school, we decided to focus on rhythm as it is one of the key aspects to work with classical violin students. So, for the rest of this work, we will assume that the improvement in the performance can be measured in terms of rhythm (See Saña (2015) for more details).

3. RECORDING SETUP

The second part of the study focuses on the designed recording sessions with students and the definition of multi-modal data to be recorded. In the last few years, the number of works related to MOCAP presented in journals and conferences have featured a notable increase. The continuously decreasing price of sensors (i.e. LeapMotion, Kinect) and acquisition platforms (i.e. Arduino or RaspberryPi), the increasing number of open developing libraries for different platforms (i.e. OpenFrameworks, Processing), in addition to the interesting results in the field of Music Information Retrieval (MIR) from different research institutions (i.e. Repovizz, SonicVisualiser and VAMP plugins), contributed to that. In this context, MOCAP data have been used for creative development (i.e. Sinyor & Wanderley (2005), Bisig & Palacio (2016) or Sarasua et al. (2016)), Human Computer Interaction (HCI) (i.e. Martin et al. (2016), Tanaka (2010) or Hemery et al. (2016)), and pedagogical applications (i.e. Hochenbaum & Kapur (2013), Chen et al. (2016) or Xiao & Ishii (2016)). Focus-

ISBN: 978-84-697-2303-6

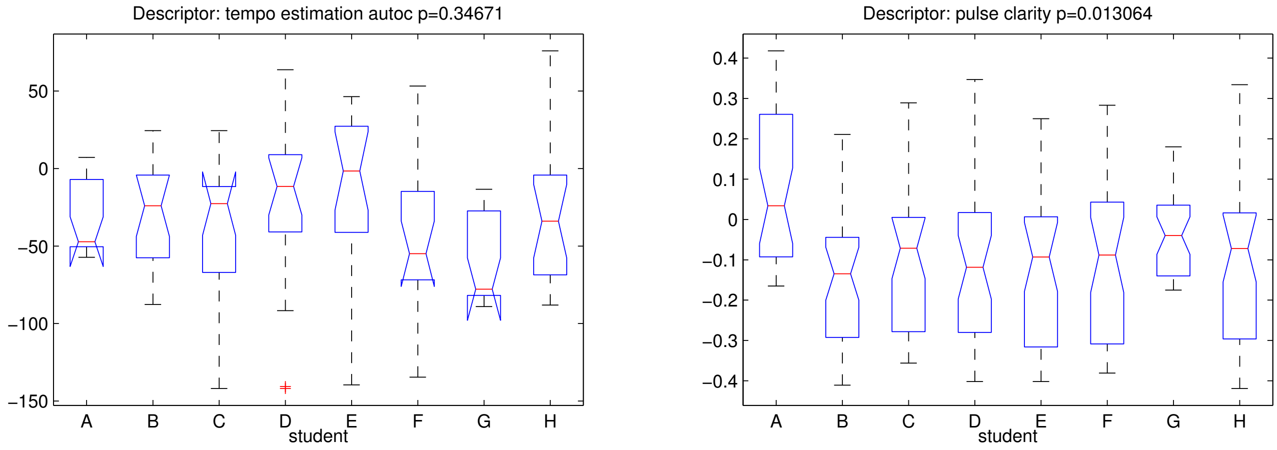


Figure 1: 1-way anova analysis plots for (a) tempo estimation (auto-correlation) descriptor on student variable, using bow-force estimation stream, and (b) pulse clarity descriptor on student variable, using pickup stream.

ing on the work here presented, a large number of works related to gesture caption and analysis from violin performances cam be found (i.e. Kimura et al. (2012), Marchini et al. (2011), Young (2009), Overholt (2005), or Young (2002)).

In parallel to the analysis of the required tools for the recording sessions, the recording procedure is debated. It consists on a set of ten recording sessions with eight students playing both classical and fiddle pieces/exercises, with some of these students following the classical tradition and others following both classical and fiddle traditions. For the preliminary study, all the students come from the Escola Superior de Música de Catalunya. For all the exercises, students and sessions, we created a multi-modal collection with video (from 3 cameras: foot, front-side and general views), audio (pickup attached to the violin and two microphones: one close to the violin and the other one far enough to include the room effects), and bow-body relative position information (from the Polhemus system developed by the Music Technology Group at Universitat Pompeu Fabra (Maestre et al., 2010)). A full description of the recording process can be found in Guaus et al. (2013).

4. ANALYSIS

4.1 Preliminary analysis

All the collected data feed the statistical analysis using state of the art techniques for audio content description in the MIR discipline (See Gouyon et al. (2008), Schedl et al. (2014) or Moffat et al. (2015) for an overview in this topic). As mentioned above, the analysis is centered in rhythmic aspects of music. Then, from the huge list of the available descriptors, this research focuses on *length*, *beatedness*, *event density*, *tempo estimation (autoc)*, *tempo estimation (spec)*, *pulse clarity*, *low energy*, *onsets*, *attack time*, and *attack slope*. All these descriptors are available in the MIRToolbox for Matlab (Lartillot & Toiviainen (2007)). All of them are normalized with respect to the descriptors obtained from recordings by expert teachers in their musical tradition (i.e. Classical or Fiddle) playing all the exercises.

The statistical influence between the exercises, students and sessions in the recorded performances are computed through an ANOVA analysis with data from descriptors derived from audio and MOCAP streams. Results reveal that Variations in the *Pulse Clarity* and *Tempo Estimation* of the audio recorded from the pickup were explained by the two groups of students described above (See Figure 1 for details). Moreover, beyond the numerical results, it is surprising to observe how it is possible to identify the two groups of students with audio recordings from data from a simple pickup, and focus the analysis on the *standard* pulse clarity and tempo estimation descriptors. In other words, for this purpose, MOCAP data is not required, and descriptors may be computed from a state of the art analysis plug-in.

4.2 Verification of results

A second benchmark of recordings and analysis is designed under these conditions (i.e. using a pickup and computing tempo estimation and pulse clarity descriptors) with students from the Conservatorio Superior de Msica Salvador Seguí de Castellón. The hypotheses we want to verify are:

H1: Students from the classical tradition with some training in the jazz/folk tradition perform rhythmically different (*Better?*) with respect with those students without this training

H2: The two groups of students can be identified with the analysis of the pulse clarity and tempo estimation from audio recordings.

A side objective of this new set of recordings and analysis is to include the research reproducibility (Vandewalle et al. (2009)) allowing small music schools to quantitatively analyze rhythmic aspects of music from their students. So, the tools we use are a small pickup attached to the violin for recordings and the descriptors are computed by using the widely extended VAMP plug-ins distributed

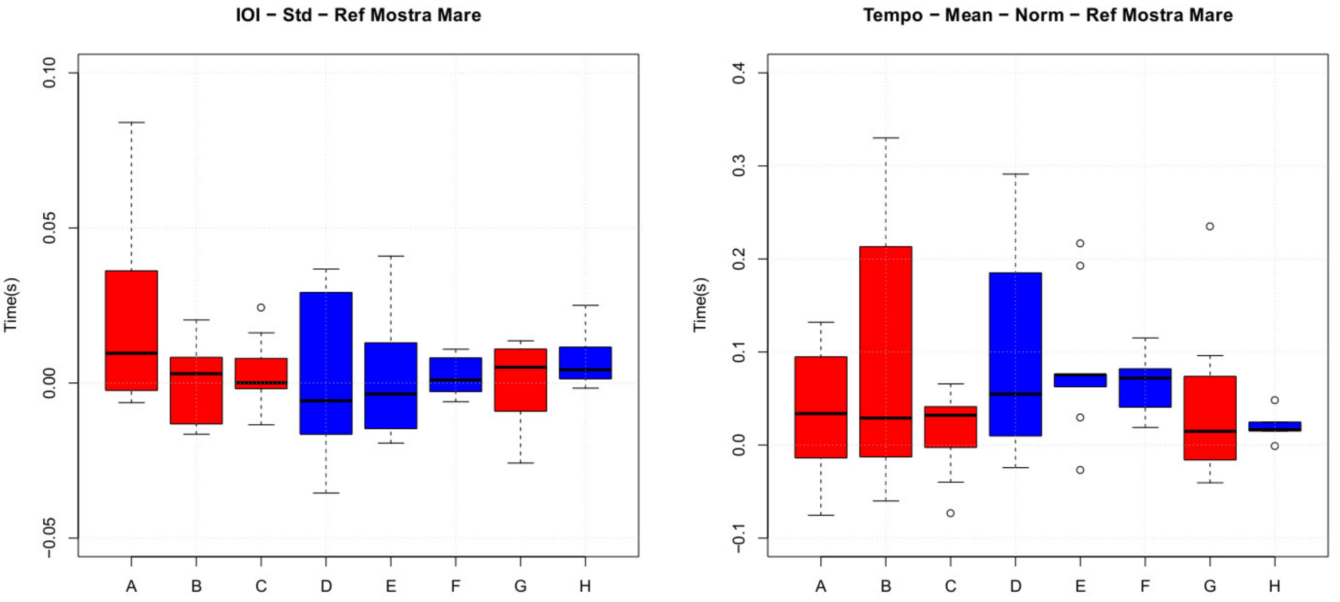


Figure 2: Results of the statistical analysis for (a) Pulse clarity and (b) beat estimation using audio data from a pickup for students for the classical tradition (blue) and students from both classical and fiddle tradition (red).

by the QMUL. Specifically, the setup and tools we use are summarized as follows:

Sonic Annotator: Sonic Annotator is a batch tool for feature extraction and annotation of audio files allowing audio analysis results to be output in a variety of formats (http://vamp-plugins.org/sonic-annotator/).

QMUL VAMP plug-ins: Sonic Annotator uses the VAMP architecture to allow researchers to implement their own algorithms for sharing research results with the whole community (http://vamp-plugins.org/plugin - doc/qm-vamp-plugins.html#qm-onsetdetector).

Specifically, by using VAMP plug-ins, we compute:

- Pulse clarity as the standard deviation of the Inter Onset Intervals (IOI) as described by Bello et al. (2005).
- Beat estimation using Tempo and Beat Tracker as described by Davies & Plumbley (2007).

Figure 2 shows the results of statistical analysis for (a) pulse clarity and (b) beat estimation from audio data for students in the classical tradition (blue) and students in both classical and fiddle tradition (red) for all the played exercises through all the sessions. Median values in pulse clarity are, in general, higher for students enrolled in both traditions at the same time than beat estimation is closer to zero. As the results are relative to the pulse clarity and beat estimation obtained by a reference teacher, we conclude that the second group of students (red) have more control in the rhythmic aspects of their performances. An this is what we wanted to demonstrate.

5. CONCLUSIONS

This paper presented a methodology for distinguishing classical tradition violin students who take lessons in the fiddle tradition with respect to those who don't. Two main conclusions can be extracted from this work. First, from the pedagogical point of view, presented results may support music schools to encourage music teachers and students the inclusion of music from multiple traditions in their daily practices and repertoire. Second, after some preliminary tests, we observed how the caption system can be implemented with a simple commercial pickup and the statistical analysis can be based on two state of the art audio descriptors available on the Internet. Due to the simplicity of the system, research reproducibility is guaranteed allowing music schools to replicate the experiments and monitor the learning process of their students.

6. ACKNOWLEDGEMENTS

The motion capture process took benefit from the technologies developed at the Music Technology Group (MTG), Universitat Pompeu Fabra (UPF), for the SIEMPRE project (EU-FP7-ICT-FET, Social Interaction and Entrainment using Music PeRformance Experimentation).

7. 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 Transactions in Speech and Audio Processing*, *13*(5), 1035–1047.

Bisig, D. & Palacio, P. (2016). Neural narratives: Dance with virtual body extensions. In *Proceedings of the 3rd Interna-*

tional Symposium on Movement and Computing, MOCO '16, (pp. 4:1–4:8)., New York, NY, USA. ACM.

Chen, L., Gibet, S., Marteau, P.-F., Marandola, F., & Wanderley, M. M. (2016). Quantitative evaluation of percussive gestures by ranking trainees versus teacher. In *Proceedings of the 3rd International Symposium on Movement and Computing*, MOCO '16, (pp. 13:1–13:8)., New York, NY, USA. ACM.

Davies, M. E. P. & Plumbley, M. D. (2007). Context-dependent beat tracking of musical audio. *Trans. Audio, Speech and Lang. Proc.*, *15*(3), 1009–1020.

Dixon, S., Goebl, W., & Widmer, G. (2002). The performance worm: Real time visualization of expression based on langner's tempo-loudness animation. In *Proceedings of the International Computer Music Conference (ICMC)*, (pp. 361–364)., Gteborg, Sweden.

Gouyon, F., Herrera, P., Gmez, E., Cano, P., Bonada, J., Loscos, A., Amatriain, X., & Serra, X. (2008). *Content Processing of Music Audio Signals*, chapter 3, (pp. 83–160). Berlin: Logos Verlag Berlin GmbH.

Guaus, E., Saña, O., & Llimona, Q. (2013). Observed differences in rhythm between performances of classical and jazz violin students. In *Proceedings of the Sound and Music Computing Conference*, Stockholm.

Hemery, E., Manitsaris, S., & Moutarde, F. (2016). A tabletop instrument for manipulation of sound morphologies with hands, fingertips and upper-body. In *Proceedings of the 3rd International Symposium on Movement and Computing*, MOCO '16, (pp. 25:1–25:8)., New York, NY, USA. ACM.

Hochenbaum, J. & Kapur, A. (2013). Toward the future practice room: Empowering musical pedagogy through hyperinstruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, (pp. 307–312)., Daejeon, Republic of Korea. Graduate School of Culture Technology, KAIST.

Kimura, M., Rasamimanana, N., Bevilacqua, F., Zamborlin, B., Schnell, N., & Fléty, E. (2012). Extracting Human Expression For Interactive Composition with the Augmented Violin. In *International Conference on New Interfaces for Musical Expression*, (pp. 1–1)., NA, France. cote interne IRCAM: Kimura12a.

Kirke, A. & Reck Miranda, E. (2009). A survey of computer systems for expressive music performance? *ACM Surveys*, *42*(1).

Lartillot, O. & Toiviainen, P. (2007). A matlab toolbox for musical feature extraction from audio. In *International Conference on Digital Audio . . .* , (pp. 1–8).

Maestre, E., Blaauw, M., Bonada, J., Guaus, E., & Perez, A. (2010). Statistical modeling of bowing control applied to violin sound synthesis. *IEEE Transactions on Audio, Speech, and Language Processing*, *18*(4), 855–871.

Marchini, M., Papiotis, P., Maestre, E., & Pérez, A. (2011). A hair ribbon deflection model for low-intrusiveness measurement of bow force in violin performance. In *New Interfaces for Musical Expression*, Oslo, Norway.

Martin, C., Gardner, H., Swift, B., & Martin, M. (2016). Intelligent agents and networked buttons improve free-improvised ensemble music-making on touch-screens. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, (pp. 2295–2306)., New York, NY, USA. ACM.

Moffat, D., Ronan, D., & Reiss, J. D. (2015). An evaluation of audio feature extraction toolboxes. In *Proc of the 18th int. Conference on Digital Audio Effects*, (pp. 1–7)., Trondheim.

Overholt, D. (2005). The overtone violin. In *Proceedings of the 2005 Conference on New Interfaces for Musical Expression*, NIME '05, (pp. 34–37)., Singapore, Singapore. National University of Singapore.

Saña, O. (2015). *I si el jazz et solucionés els problemes?* PhD thesis, Universitat Autònoma de Barcelona, The address of the publisher.

Sarasua, A., Caramiaux, B., & Tanaka, A. (2016). Machine learning of personal gesture variation in music conducting. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, (pp. 3428–3432)., New York, NY, USA. ACM.

Saunders, C., Hardoon, D., Shawe-taylor, J., & Gerhard, W. (2004). Using string kernels to identify famous performers from their playing style. In *Proceedings of the 15th European Conference on Machine Learning (ECML)*.

Schedl, M., Gmez, E., & Urbano, J. (2014). Music information retrieval: Recent developments and applications. *Foundations and Trends in Information Retrieval*, *8*(2-3), 127–261.

Sinyor, E. & Wanderley, M. M. (2005). Gyrotyre : A dynamic hand-held computer-music controller based on a spinning wheel. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, (pp. 42–45)., Vancouver, BC, Canada.

Tanaka, A. (2010). Mapping out instruments, affordances, and mobiles. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, (pp. 88–93)., Sydney, Australia.

Vandewalle, P., Kovacevic, J., & Vetterli, M. (2009). Reproducible research in signal processing. *IEEE Signal Processing Magazine*, *26*(3), 37–47.

Widmer, G. & Goebl, W. (2004). Computational models of expressive music performance: The state of the art. *Journal of New Music Research*, *33*(3), 203–216.

Xiao, X. & Ishii, H. (2016). Inspect, embody, invent: A design framework for music learning and beyond. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, (pp. 5397–5408)., New York, NY, USA. ACM.

Young, D. (2002). The hyperbow controller: Real-time dynamics measurement of violin performance. In *Proceedings of the 2002 Conference on New Interfaces for Musical Expression*, NIME '02, (pp. 1–6)., Singapore, Singapore. National University of Singapore.

Young, D. (2009). Classification of common violin bowing techniques using gesture data from a playable measurement system. In *Proc. of the NIME Conference*.

Poster session 1

MELODIC PATTERN CROSS-OCCURRENCES BETWEEN GUITAR FALSETAS AND SINGING VOICE IN FLAMENCO MUSIC

Inmaculada Morales
University of Seville
inmaguitar@gmail.com

Nadine Kroher
University of Seville
nkroher@us.es

José-Miguel Díaz-Báñez
University of Seville
dbanez@us.es

1. INTRODUCTION

Flamenco is a rich oral music tradition with strong improvisational character from the Southern Spanish province of Andalucía. Having evolved from a singing tradition (Gamboa, 2005), the singing voice, refered to as *cante*, remains the central element in the genre's current form, accompanied by guitar playing and rhythmical hand clapping. During a flamenco performance, accompanied singing sections alternate with instrumental interludes, in which guitarists often step out of their accompanying role and perform so-called *falsetas*. In Núñez & Gamboa (2007) the term *falseta* is described as "the interpretation of a small composition with autonomous musical identity". By playing *falsetas*, guitarists contribute a piece of their own inspiration, either composed by themselves or re-interpreted from masters of the genre, to the performance as a whole. In this study, we focus on the mutual interaction between *falseta* and *cante* with respect to melodic content. We identify a large number of cases where a characteristic melodic pattern can be found in guitar *falsetas* as well as in the singing voice melody. These cross-occurrences are not limited to the same performance, but can span across decades and even genres. Based on a corpus of 50 such melodic cross-occurrences, we study their characteristics and computationally assess the melodic similarity of the detected examples. This study opens a new research line in computational ethnomusicology which can reveal novel aspects of the creation and evolution of flamenco music and furthermore gives rise to a number of technological challenges.

2. CORPUS STUDY OF MELODIC CROSS-OCCURRENCES

We gathered a representative corpus of 50 examples of melodic patterns encountered in commercial music recordings, which occur in both, a singing voice and a *falseta* melody. In 67% of the cases, the melodic fragment first occurred in a sung melody and has later found re-use in a *falseta*. In some cases, this recreation occurs either during the same performance in a call-response manner. In other cases, the *falseta* melody is taken from popular genres (i.e. *coplas* or *cuplés*) or related flamenco fusion genres, including the *rumba catalana*, flamenco rock and flamenco inspired pop music commonly referred to as "new flamenco". In the remaining 33% of the examples, a melody which first occurred in a *falseta* has later been reinterpreted in the *cante*. In some cases, the respective melodic fragment even takes on a fundamental structural role in a flamenco song. We furthermore discovered that both cases, re-use through the guitar and through the singing voice, show a tendency to take place in particular flamenco styles (specifically *bulerías* and *tangos*) and that certain guitarists show to be particularly involved in this process, most prominently *Paco de Lucía*.

3. QUANTITATE ASSESSMENT OF MELODIC SIMILARITY

Given the expressive and improvisational nature of flamenco music, occurrences of the same melodic pattern will inevitably exhibit difference by means of melodic variation and ornamentation. In order to objectively assess the similarity between two instances of the same fragment, we apply a computational melodic similarity measure. We manually transcribe the guitar *falseta* (a) and the respective sung melody segment (b) to MIDI format. For a given pair a and b we compute the *earth mover's distance* (Typke et al., 2003) $d_{a,b}$ and compute the ratio $r = \frac{d_{a,rand}}{d_{a,b}}$ with respect to the average distance of a to 500 randomly melodic fragments extracted from the *Corpus COFLA* (Kroher et al., 2016).

4. CASE STUDIES

We conducted a number of case studies, where we analyse relevant examples of the research corpus in detail with respect to the context of origin and re-interpretation of the melodic material and the amount of variation among them. One example is a melodic pattern appearing in both *falseta* and *cante* of the song *Tangos de la Sultana* recorded by singer *Camaron* together with guitarist *Tomatito* (Figure 1) in 1979. Within the song, the pattern first occurs in the guitar before it is interpreted by the singer. It furthermore forms an essential part of the vocal melody, which is repeated throughout the song. A very similar pattern is encountered in a song of the same style titled *La que quiera madroños vaya a la sierra* recorded by singer *La Repompa* in 1958. A computational analysis shows that the computed similarity between the *falseta* and the vocal section in *Tangos de la Sultana* is nearly identical to the similarity computed between the *falseta* and the respective section in *La que quiera madroños vaya a la sierra*.

ISBN: 978-84-697-2303-6

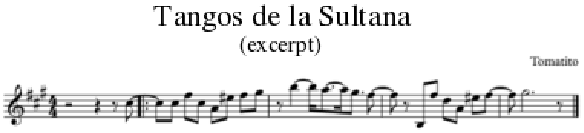


Figure 1: Manual transcription of a melodic pattern from the song *Tangos de la Sultana* recorded by *Camaron* and *Paco de Lucía*.

5. ACKNOWLEDGEMENTS

This research has received funding from the projects COFLA2 (Junta de Andalucía, P12-TIC-1362), GALGO (Spanish Ministry of Economy and Competitiveness, MTM2016-76272-R AEI/FEDER,UE) and CONNECT (EU-H2020/MSCA under grant agreement 2016-734922).

6. REFERENCES

Gamboa, J. M. (2005). *Una historia del flamenco*. Espasa Calpe.

Kroher, N., Díaz-Báñez, J.-M., Mora, J., & Gómez, E. (2016). Corpus cofla: a research corpus for the computational study of flamenco music. *Journal on Computing and Cultural Heritage (JOCCH)*, *9*(2), 10.

Núñez, F. & Gamboa, J. (2007). *Flamenco de la A a la Z.* S.L.U. Espasa libros.

Typke, R., Giannopoulos, P., Veltkamp, R. C., Wiering, F., & Van Oostrum, R. (2003). Using transportation distances for measuring melodic similarity. In *Proceedings of the International Music Information Retrieval Society (ISMIR)*. Johns Hopkins University.

EXTRACTION AND CLASSIFICATION OF ORNAMENTATION IN FLAMENCO SINGING: AN EVOLUTION-BASED APPROACH

Inmaculada Marqués, Nadine Kroher, Joaquín Mora, José-Miguel Díaz-Báñez
University of Seville
inma.marques.donaire@gmail.com, nkroher@us.es, mora@us.es, dbanez@us.es

1. INTRODUCTION

In music traditions around the world, melodic ornamentation is used by performers as an expressive resource to embellish and add individual interpretation to a melody. In flamenco singing, *cante flamenco*, it is precisely the ornamentation which defines the flamenco aesthetics and without it the *cante* would not be flamenco.

Each flamenco style is characterized by a distinct prototypical melody (melodic skeleton), which can be subject to a great range of ornamentation and variation (Mora et al. (2016)). Despite the commonly referenced presence of ornamentation in flamenco music, only few systematic approaches have studied this phenomenon. To this day, there does not exist any established taxonomy of ornaments and the structural importance of melisma remains unexplored. Gómez et al. (2011) proposed a computational approach to recognition and characterization of flamenco ornamentation. A set of pre-defined ornament types, mainly borrowed or adapted from classical music theory, is extracted from a corpus using a the Smith-Waterman algorithm (Smith & Waterman (1981)). Other approaches to recognition and characterization of ornamentation have been proposed in the context of popular ornamentation, see for example Puiggròs et al. (2006); Perez et al. (2008); Giraldo & Ramírez (2016). In this work we propose a new computational strategy to detect and characterize ornamentation in flamenco singing.

2. METHODOLOGY

We approach the problem of extracting and characterizing ornamentation by focusing on a specific type of flamenco *cantes* which have evolved from traditional popular chants. Flamenco singers extend the popular melody by introducing melodic ornamentation and variation, mainly in the form of melismatic ornamentation. Consequently, by comparing flamenco performances to popular version of the same melody, we can quantitatively assess, extract and characterize the ornamentation introduced by the flamenco artist.

The outline of the approach is as follows. We process a corpus containing recordings of the popular melody and various flamenco interpretations of the same chant. For each recording, we first perform a computer-assisted singing voice transcription using the CANTE (Kroher & Gómez, 2016) software. We then use the gap-tolerant Needleman-Wunsch (Needleman & Wunsch, 1970) alignment algorithm to align the flamenco performance transcriptions to the popular melody. Assuming that unmatched notes in flamenco performances correspond to added content with respect to the popular melody, we can isolate sections of the song where ornamentation occurs. An example is shown in Figure 1 (top), where notes marked in red correspond to the popular melody and the black segments are the extracted ornaments. We use the isolated ornaments to establish a taxonomy of typical ornamentation in flamenco music, borrowing and extending concepts defined in the context of classical and medieval music. More specifically, we represent the isolated the ornaments in *Parsons* code (Parsons (1975)) as a concatenation of basic *neums*. In this way, we can analyze the occurrence of small melodic atoms and their combinations.

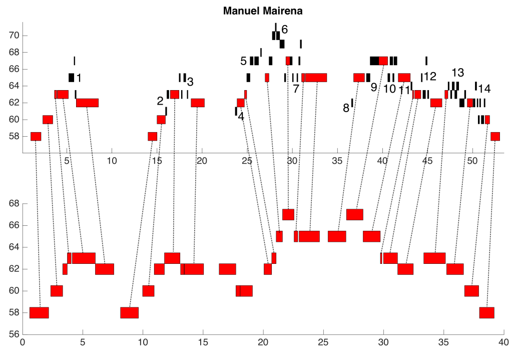


Figure 1: The "differences" in the melodic contour contains the ornamental resources. Bottom: the popular version. Top: the version performed by flamenco singer *Manuel Mairena*.

3. CASE STUDIES IN RELIGIOUS CONTEXT

In this pilot study we consider the case of a religious chant ("Santo Dios") which is performed in a social-religious context of Mairena del Alcor (Seville, Spain) and has evolved into a flamenco style. For this case, we collected several versions (including the popular and flamenco versions) in order to find the ornaments representing the formal change to the flamenco form.

This method is proposed in Marqués et al. (2012) to study the process of flamenco evolution. The case of ("Santo Dios") is considered as a live model to investigate the cultural preferences that influence the creation and evolution of flamenco music.

ISBN: 978-84-697-2303-6

4. ACKNOWLEDGEMENTS

This research has received funding from the projects COFLA2 (Junta de Andalucía, P12-TIC-1362), GALGO (Spanish Ministry of Economy and Competitiveness, MTM2016-76272-R AEI/FEDER,UE) and CONNECT (EU-H2020/MSCA under grant agreement 2016-734922).

5. REFERENCES

Giraldo, S. & Ramírez, R. (2016). A machine learning approach to ornamentation modeling and synthesis in jazz guitar. *Journal of Mathematics and Music*, *10*(2), 107–126.

Gómez, F., Pikrakis, A., Mora, J., Díaz-Báñez, J. M., Gómez, E., & Escobar, F. (2011). Automatic detection of ornamentation in flamenco. In *Fourth International Workshop on Machine Learning and Music MML.*

Kroher, N. & Gómez, E. (2016). Automatic transcription of flamenco singing from polyphonic music recordings. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *24*(5), 901–913.

Marqués, I., Díaz-Báñez, J. M., & Mora, J. (2012). El canto (cante) al cristo de la cárcel en mairena del alcor. In *Boundaries between Genres: Flamenco and Others Musical Oral Traditions. Proc. FMA 2012*, (pp. 51–60).

Mora, J., Gómez, F., Gómez, E., & Díaz-Báñez, J. M. (2016). Melodic contour and mid-level global features applied to the analysis of flamenco cantes. *Journal of New Music Research*, *45*(2), 145–159.

Needleman, S. B. & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, *48*(3), 443–453.

Parsons, D. (1975). *The directory of tunes and musical themes*. Cambridge, Eng.: S. Brown.

Perez, A., Maestre, E., Kersten, S., & Ramirez, R. (2008). Expressive irish fiddle performance model informed with bowing. In *Proceedings of the International Computer Music Conference (ICMC 2008).*

Puiggròs, M., Gómez, E., Ramírez, R., Serra, X., & Bresin, R. (2006). Automatic characterization of ornamentation from bassoon recordings for expressive synthesis. In *Proceedings of International Conference on Music Perception and Cognition*, (pp. 22–26).

Smith, T. F. & Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of molecular biology*, *147*(1), 195–197.

THE AEPEM COLLECTION: A SET OF ANNOTATED TRADITIONAL FRENCH MUSIC SCORES

Pierre Beauguitte
Dublin Institute of Technology, School of Computing
pierre.beauguitte@mydit.ie

ABSTRACT

The aim of this paper is to present the AEPEM collection, consisting of more than five thousand scores of French traditional melodies. The original material and the digitized collection are described. A short statistical analysis is performed to compare this collection to existing ones in terms of melodic profiles.

1. INTRODUCTION

The AEPEM is an association, created in 2004, working on French traditional music. Its name is an acronym for what can be translated as "association for studying, promoting, and teaching traditional music from French provinces". Aside from its role as a record label, the main focus of the association has been the publication of a digital music library.

During the second half of the 19th century and the beginning of the 20th, popular music and songs have been collected in many parts of France. The collectors, or "folklorists", in charge of this work have documented the artifacts with various degrees of precision. Books containing the scores, sometimes alongside other ethnographic considerations, have been published at that time. Many of these books are publicly available via the French National Library [1] or on the website `archive.org`.

The AEPEM is aiming at making these music scores available, in a digital format, and in a single repository.

2. THE DIGITAL LIBRARY

The scores have been manually digitized using the software Melody Assistant. [2] The files created are in a proprietary format, with a `.myr` extension. MIDI and ABC files were automatically generated, and are also available. 5418 melodies are published at the time of writing, but this number is growing as more books are being digitized.

When it is available in the original book, the following metadata is given:

- title
- incipit: first line of the lyrics
- type of melody: dance tune, lullaby...
- location and date of collection
- name of the singer
- location and date of birth of the singer
- name of the collector

3. COMPARATIVE ANALYSIS

In this section, we compare the AEPEM collection with:

- the Meertens Tune Collection - Large Corpus (MTC-LC), presented in van Kranenburg et al. (2014), and containing 4830 Dutch songs
- O'Neill's collection *The Dance Music of Ireland*, containing 1001 Irish traditional tunes [3] (O'Neill (1907))

First, we simply count the occurrences of different intervals in all melodies of the corpora. The bar charts in Figure 1 shows the relative frequency of all intervals. In AEPEM as in MTC-LC, ascending and descending major seconds, and unison, are the most common melodic intervals. Unison occurs much less frequently in O'Neill's collection.

Second, we count the frequencies of pairs of successive intervals, that give a richer description of the melodic contours. The heatmaps in Figure 2 show these frequencies, restricted to intervals between descending and ascending fifths. The X- and Y-axis represent the first and second interval of the pair, respectively. A striking resemblance appears between the AEPEM and the MTC-LC heatmaps, but is not shared with O'Neill's.

Further analysis could be conducted using n-grams of intervals, or other sets of features, to reveal common characteristics and specificities of the different corpora. More importantly, musicological and perceptual analysis could be conducted to assess whether or not these objective measurements correlate with perceived similarities.

4. CONCLUSION

We have introduced the AEPEM collection, and described both its sources and the digitized collection, available on request. [4] A short statistical analysis revealed similarities of the melodic contours in this collection and in MTC-LC.

The AEPEM collection is the result of a collaborative effort started in 2004. We believe it can be a valuable resource for the study of French traditional music, from the perspective of ethnomusicology as well as computational analysis. Thanks to the metadata provided, and the availability of scanned versions of many of the books, tasks such as geographical clustering or optical music recognition can be tackled.

[1] `gallica.bnf.fr`

[2] `www.myriad-online.com/en/products/melody.htm`

[3] ABC transcriptions available at `trillian.mit.edu/~jc/music/book/oneills/1001/`

[4] `www.aepem.com/contact`

ISBN: 978-84-697-2303-6

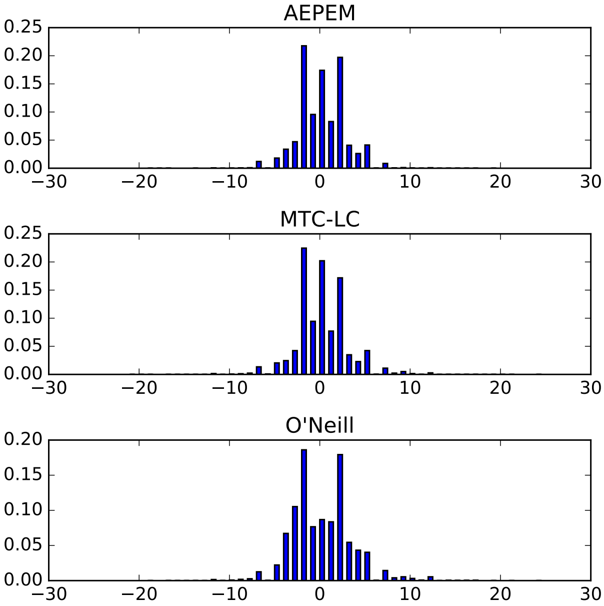


Figure 1: Relative frequency of intervals (X-axis in semitones)

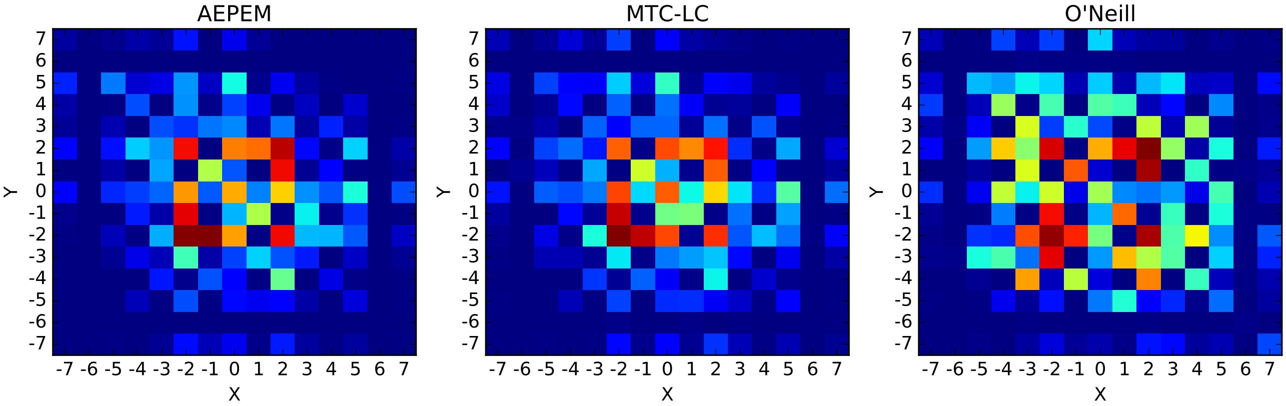


Figure 2: Heatmaps of the relative frequency of pairs of successive intervals (axes in semitones)

5. REFERENCES

O’Neill, F. (1907). *The Dance Music of Ireland*. Chicago: Lyon and Healy.

van Kranenburg, P., de Bruin, P. . M., Grijp, L. P., & Wiering, F. (2014). The Meertens Tune Collections. Meertens Online Reports 1, Amsterdam: Meertens Institute.

DEMONSTRATION OF CHORD DETECTION AND ANALYSIS SOFTWARE

Eva Ferková, prof. PhD.
Department of Music Theory, Academy of Performing Arts, Zochova 1, Bratislava, Slovakia
ferkova@vsmu.sk

Michal Šukola
Accenture, s.r.o, Plynárenská 7/C, Bratislava, Slovakia
michal.sukola@gmail.com

1. INTRODUCTION

The demonstration of software is connected with the contribution focused on ethno-musicological and harmonic analysis of adaptations of Slovak folk songs by Slovak-American composer Miloslav Francisci. The software is oriented on detection of sounds, which are built up with thirds, that means with distances of 3 or 4 semitones modulo 12 between each sounding pitch. If there is such vertical structure detected, the chord is named according to the attached table. The presented software improves the Harmanal tool, which was presented to analyse e.g. Schubert´s and Mozart´s compositions in the years 2007-2009.

2. INPUT DATA

Input data are in MIDI format. MIDI is the most widespread format for digitalization of music information. There are two most severe problems in work with this type of music information: 1. the same digit for enharmonically different notes effects ambiguity in detection of concrete tonal-key affiliation. 2. the low quality of MIDI file of majority of compositions downloaded from internet from the point of view of time (duration of each pitch). MIDI files on Internet are previously created by playing on MIDI keyboards by people, who are not able to play machine-like time-precisely. Therefore some sounds are in these MIDI files included also to previous beat or following one, where it should not exactly sound (the composer didn´t write it so). The mixture of sounds belonging to neighboring beats causes difficulties in chord-detection.

3. CHORD AND MELODIC TONES DETECTION

The procedure of chord tones detection is based on the looking for distance of 3 or 4 between MIDI numbers for each pitch-pair after shifting there to the nearest position each other. Those pitches, which are not used in this distance, are evaluated as melodic ones, and are excluded from further calculation. If there is no structure of thirds detected in one beat, it is included into another beat for looking for arpeggiated chord till there is another barline.

4. NAMES OF DETECTED CHORDS

Names of chords are easy to understand, while they are tending to be international. Signs for triads are “+” for major triad, “-“ for minor triad, etc. The inversion of triad is signed by numbers as usual in basso continuo. The inversions of seventh chords are signed by numbers – fractions, which sign the interval between bass-pitch to seventh and to root of chord (5/6, ¾ and 2 – as ½).

5. OUTPUT, RESULTS AND THEIR USAGE

5.1 Form of output

The output is visualized in Sibelius score file, the signs of chords are located under the bottom staff, horizontally under the particular beat. If the chord is repeated, the repetition sign is “=”.

5.2 Results – their significance and limitations

Results might offer a way to define one of style-features of musical thinking of analyzed collection of songs (compositions). Current version of software detects only structure of chords, not their tonal functions as tonic or dominant. For this possibility we are currently working on another extension – automatic tonal-key detection. Then it would be possible to allocate the harmonic function for every chord and eventually to find and define the cadential progressions and harmonic dynamism.

6. REFERENCES

Ferková, E. – Urbancová, H. (2017). Adaptations of Slovak Folk Songs for Piano in the Context of (Ethno) Musicological Analysis. Malaga *FMA 2017 Proceedings*

Ferková, E. (2009). Computer-Aided Investigation of Chord Vocabularies: Statistical Fingerprints of Mozart and Schubert in: *Mathematics and Computation in Music. Communications in computer and information science, 2009, Volume 37, I, Part 8,* pp. 250-256

Ferková, E. (2007). Chordal Evaluation in MIDI-Based Harmonic Analysis: Mozart Shubert and Brahms. In: *Computing in Musicology 15. Tonal Theory for the Digital Age* Stanford CCARH pp. 186-200

Selfridge-Field, E. (1997). Beyond MIDI. Cambridge, Massachusetts, London The MIT Press

Piston, W. **(**1987). Harmony. Revised by DeVoto. M. New York-London: W. W. Norton & Company

ISBN: 978-84-697-2303-6

Structure in number of semitones	English name of the chord	Sign of chord
4-3	Major triad	+
3-4	Minor triad	-
4-4	Augmented triad	++
3-3	Diminished triad	--
4-3-3	Dominant seventh	D7
4-3-4	Major seventh	Maj7
3-4-3	Minor seventh	Min7
3-3-3	Diminished seventh	Dim7
3-3-4	Half-diminished seventh	Dm7
4-4-3	Augmented seventh	Aug7
3-4-4	Minor-major seventh	Min+7

Table 1. Detected chords, their names and signs. From the left: interval structure of the chord in basic position in number of semitones between neighboring tones from the root, English name of the chord, used sign.

Chord duration (%)	**Chord types**											
File	**Maj 5**	**Min 5**	**Dim 5**	**Aug 5**	**D7**	**Dim 7**	**Dm 7**	**Maj 7**	**Min 7**	**MinMaj 7**	**Aug 7**	**No**
05_Tou_nasou_dolineckou.mid	47%	1%	0%	0%	35%	0%	0%	0%	0%	0%	0%	17%
07_Este_sa_nevydam.mid	38%	25%	0%	0%	23%	0%	0%	0%	0%	0%	0%	8%
19_Ovce_moje,_ovce_problem.mid	14%	31%	0%	0%	20%	0%	19%	3%	5%	1%	0%	6%
72_Nichto_nezna,_nebude_znac.mid	74%	4%	2%	0%	11%	0%	2%	0%	0%	0%	0%	6%
73_Pri_Presporku_verbuju.mid	34%	24%	5%	0%	15%	9%	0%	0%	0%	2%	0%	11%
74_Zahradka,_zahradka.mid	50%	10%	3%	0%	30%	0%	0%	0%	0%	0%	0%	8%
75_V_tej_nasej_zahradke.mid	53%	10%	3%	0%	18%	0%	0%	0%	0%	0%	0%	18%
76_Leti,_leti_roj.mid	20%	30%	3%	0%	18%	0%	2%	8%	11%	0%	0%	7%
78_Ide_suhaj_po_dvore.mid	22%	36%	3%	0%	33%	3%	0%	0%	0%	0%	0%	3%
90_Kopala_studienku.mid	23%	35%	3%	0%	33%	1%	4%	0%	1%	1%	0%	0%
Total duration in all files	36%	20%	2%	0%	26%	1%	4%	1%	2%	0%	0%	8%

Figure 1. The table of statistical results of appearance of chords, detected by automated computerized chord analysis

Poster session 2

Folk Music Analysis
MÁLAGA, Spain 14-16 June 2017

Adaptations of Slovak Folk Songs for Piano in the Context of (Ethno)Musicological Analysis

Eva Ferková
Department of Music Theory, Academy of Performing Arts, Bratislava, Slovakia
ferkova@vsmu.sk

Hana Urbancová
Institute of Musicology, Slovak Academy of Sciences, Bratislava, Slovakia
Hana.Urbancova@savba.sk

1. Introduction

This paper presents the results of collaboration between an ethnomusicologist and a music theoretician, analysing adaptations of folk songs by a composer of the late 19th century.

In the introduction we present the Slovak-American composer Miloslav Francisci (1854 –1926), whose adaptations of Slovak folk songs, designed for concert performance, contributed to the propagation of Slovak folk music in America. His two-part cycle of folk songs adapted for piano, entitled *Trávnice* ([Haymaking Songs], 1892, 1893) contains 200 piano miniatures in total. Francisci took models for elaboration from *Slovenské spevy* ([Slovak Songs], 1880 – 1926), a collection which developed from the largest organised initiative of collecting in 19th century Slovakia. In the history of Slovak music these adaptations demonstrate the transition from simple harmonisations with a practical function to the individual approach of the composer, adapting the folklore model with emphasis on the aesthetic function.

2. Ethnomusicological point of view

The first part of the paper explains the results of an ethnomusicological analysis of the folk tunes which served the composer as models for adaptation. Based on the analytic system of the Slovak ethnomusicological school (J. Kresánek, A. Elscheková – O. Elschek), the style layers of Slovak folk music are defined and their representation in the song repertoire chosen by the composer is identified. We focus particularly on some specific features of the folk tunes, which we will further confront with the compositional approach of Miroslav Francisci.

3. Analytical point of view

The second part of the paper presents the results of a harmonic analysis of the piano miniatures, with the aim of revealing the level of artistic input by the composer. In harmonising all types of songs Francisci attempted to use cadence progressions and functional harmony based on major and minor keys, while also employing more advanced relationships such as double dominants, modulations, deceptive cadences, chromaticisations and alterations, etc.

In the analytic section the paper compares the results of a traditional analysis of harmonisations, which we present in a table listing all of the songs. The table includes names of songs, tonal keys of harmonisations, and harmonic points of interest connected with the use of tonal cadence progressions.

4. Preparation for future research of computer determination of chords and tonal keys

The results presented there will be compared with a partial computerised (automatised) detection of 11 basic types of chords (triads and seventh chords) in a number of songs. We also concisely introduce the idea of a computerised tonal analysis (in preparation), which could offer an automatised (computer) determination of major or minor key not only in these songs but in major-minor harmonised folk songs of any kind, including from other regions.

5. REFERENCES

Elscheková, A. (1978), Stilbegriff und Stilschichten in der slowakischen Volksmusik. In: Studia Musicologica, Vol. 20, p. 263-303.

Elscheková, A. – Elschek, O. (2005), Úvod do štúdia slovenskej ľudovej hudby. Bratislava: Hudobné centrum.

Filip, M. (2012), Vývinové zákonitosti klasickej harmónie. Bratislava: Hudobné centrum.

Francisci, M. (1892). Trávnice. 100 Slovenských národných piesní. Sväzok I. Turčiansky Svätý Martin: Knihtlačiarsky účastinársky spolok.

Francisci, M. (1893). Trávnice. 100 Slovenských národných piesní. Sväzok II. Turčiansky Svätý Martin: Knihtlačiarsky účastinársky spolok

Kresánek, J. (1997), Slovenská ľudová pieseň zo stanoviska hudobného. Bratislava: Národné hudobné centrum.

ISBN: 978-84-697-2303-6

Song n.	Song name	Key	Specialty	Form
72	Nichto nezná, nebude znac	G major	Diatonic melody, cadential harmony, unfinished cadence in e minor, closure on dominant	ab
73	Pri Prešporku verbujú	D minor -g minor	Rubato, ornamentation, triplets, tonal uncertainty, deceptive cadence in d minor.	ab (9b)
74	Záhradka, záhradka	G major	Diatonic melody, ornamentation, 4 bars upper voice melody, 4 bars lower voice melody. Cadence with II. degree as subdominant	ab (10b)
75	V tej našej záhradke	D major-A major-D major	Simple cadential harmony, arpeggiated chords	ab
76	Letí, letí roj	C minor-E flat major-c minor	Simple cadential harmony, using Tonic and Dominant, in last cadence II34 as Subdominant. Scale movement	ab
77	Zaviau vetrík cez dolinu	a minor-C major-a minor	Diatonic melody, cadence with altered chord in C major key g-b-d sharp-f. Quick scale movement (4x)	a-ba
78	Ide šuhaj po dvore	b flat minor-A flat major-D flat major-b flat minor	Modulating harmony, abnormally written chord of diminished 7 to Dominant in F minor as e- g - b flat - d flat - f flat.	aa1a2 (9b)

Figure 1. Segment of table of all 200 songs and their basic analytical information

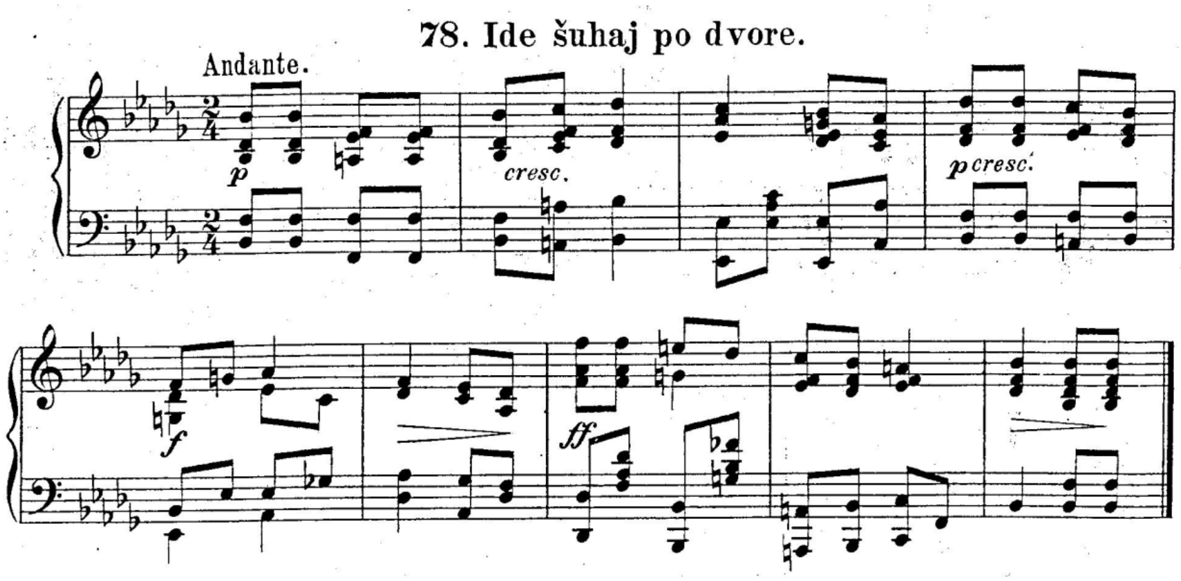


Figure 2: example of printed score of one folk song, adapted for piano by Miloslav Francisci. Trávnice I.

CROSSED-EYED CHORO: FORMAL DEFORMATIONS IN BRAZILIAN CHORO

Cibele Palopoli
University of São Paulo
cibele.palopoli@gmail.com

1. INTRODUCTION

The Portuguese Court arrived in Brazil in 1808. Among other news, they brought to the New World several European salon dances, such as polka, schottish, waltz, mazurka, quadrille, and redowa. Over the years, the last three fell into disuse, while the waltz, the schottish, and especially the polka became each time more popular. The polka was so strongly incorporated that some historians even considered it as a Brazilian creation (Pinto, 2014 [1936]).

Choro was originally born from a Brazilianized way to interpret this European repertoire by ensembles formed by flute, *cavaquinho*, and guitar. Its emergence dates back to the 1870s (Kiefer, 1979: 23), while its consolidation as a genre returns to the 1910s (Cazes, 2005 [1998]: 19)[1].

The repertoire of Brazilian composers of the mid-nineteenth century such as Joaquim Callado, Chiquinha Gonzaga, Ernesto Nazareth, and Anacleto de Medeiros, includes polkas, schottisches, and waltzes, besides other new Brazilian creations, such as *maxixe*, Brazilian tango, and the homonym, choro[2]. At this point, music was not necessarily to be danced anymore, but to be contemplated.

Among other aspects, Choro inherited from its European ancestors[3] the tripartite structure in Rondo form (ABACA) with immediate repetition of each new section (therefore, AABBACCA), and contrasting tonalities between each part (see Tables 1 and 2).

When in duple meters, each part has 16 bars divided in two equal halves (antecedent, which leads to the Dominant, and consequent, which leads to the Tonic). When in triple meters (waltzes), this structure is doubled to 32 bars in each section.

For several years this structure was strictly respected. Pixinguinha, one of the Choro's icons, took eleven years to publish his masterpiece, *Carinhoso* (1917), once it is formed only by two sections, having the B part 24 bars. Nevertheless, this composition preserves phrases multiple of eight bars.

Late Choro compositions, such as those written by another Choro's master, Jacob Bittencourt, abandoned the C part, but carried on the preservation of symmetric phrases on both A and B parts.

At the turn of the twentieth and the twenty-first centuries the hence called "deformation" of the original Rondo form became each time more usual. The main objectives of this study are to map the route of these formal deformations until the Choro produced nowadays and to verify if these new compositions could still been considered as Choros.

The title of this paper, *Crossed-Eyed Choro* [*Choro vesgo*] is inspired by a piece made by Zé Barbeiro (b. 1952), a Choro composer and seven strings guitar player based in São Paulo, who has currently modifying Choro's traditional conceptions among his more than 220 works.

2. METHODS

The method employed in this paper is the statistical approach based on Choro canons composed from 1910 to 2015. The wide range of the chronological period is justified because only a reduced number of compositions from the mentioned period are not structured on the Rondo form. Formal music analyses of this repertoire were made based on titles directed to the study of structural aspects on Classical Music (Berry, 1966; Caplin, 1998; Mathes, 2007), and on Choro (Almada, 2006, 2012).

3. RESULTS

The results show an increasing number of formal deformations over the years with the use of complex compositional procedures, such as asymmetric and mixed metrics, prolongations or ruptures in motifs and cadenzas etc.

4. DISCUSSION

What are the causes of the abovementioned deformations? Is it derived from the contact with other musical genres? Could it indicate the desire of dissociating Choro with its European roots?

5. CONCLUSIONS

In conclusion, I reflect about the arising of a new stylistic school of Choro, which has been orally called as Contemporary Choro.

[1] According to William Hanks, "when viewed as an interaction between social constructs and musical content, genre may be seen to offer: 1) a framework that a listener may use by which to orient themselves; 2) procedures to interpret the music; and 3) a set of expectations" (Hanks apud Beard & Gloag, 2005: 72).

[2] All of these genres are part of the musical and cultural manifestation called **C**horo. I therefore distinguish the specific genre choro with the lowercase initial.

[3] An African ancestry could also be verified especially concerning some rhythmic patterns, but it would be an issue for another paper.

ISBN: 978-84-697-2303-6

		Part A	Part B	Part C
Harmonic relationships possibilities	Major Keys	I	vi V	IV
	Minor Keys	i	III I	I III VI

Table 1. Some harmonic relationships possibilities in three-part Choros.

		Part A	Part B
Harmonic relationships possibilities	Major Keys	I	vi IV III
	Minor Keys	i	III I VI

Table 2. Some harmonic relationships possibilities in two-part Choros.

6. REFERENCES

Almada, C. (2006). *A estrutura do Choro:* com aplicações na improvisação e no arranjo [The Structure of Choro: with Applications in Improvisation and Arrangement]. Rio de Janeiro: Da Fonseca.

Almada, C. (2012). O choro como modelo arquetípico da Teoria Gerativa da Música Tonal [Choro as an Archetypal Model of the Generative Theory of Tonal Music]. *Revista Brasileira de Música*, *25*(1), 61-78.

Beard, D. & Gloag, K. (2005). *Musicology:* The Key Concepts. New York: Routledge.

Berry, W. (1966). *Form in Music.* Englewood Cliffs: Prentice-Hall.

Caplin, W. (1998). *Classical Form.* USA: Oxford University Press.

Cazes, H. (2005 [1998]). *Choro:* do quintal ao Municipal [Choro: from the Backyard to the Municipal Theater]. 3rd. edition. São Paulo: Editora 34.

Kiefer, B. (1979). *Música e dança popular:* sua influência na música erudita. [Popular Music and Dance: its Influence in Classical Music]. Porto Alegre: Editora Movimento.

Mathes, J. (2007). *The Analysis of Musical Form.* Upper Saddle River: Prentice Hall.

Pinto, A. G. (2014 [1936]). *Choro:* reminiscências dos chorões antigos [Choro: Reminiscences of Old Musicians]. 3rd edition. Rio de Janeiro: Acari Records, 2014 [1936].

METER IN “ESPERANDO NA JANELA” (2000): A GLIMPSE INTO HYPERMETRIC SHIFTS AND PERCEPTION

Eduardo Solá Chagas Lima
Andrews University
solachagas@andrews.edu

1. INTRODUCTION

This abstract entails a brief metrical analysis of “Esperando na Janela” (2000), a sample of *Forró* music, which lies at the heart of Brazilian northeastern folklore. In this research, I explore this song’s meter and hypermetric structure, with special attention to how its text and musical clothing allow for a metrical shift, thus requiring hypermetric reinterpretation on the part of the listener. I will also discuss potential impacts of this shift on *forró* dance. This will finally lead to a brief investigation of the perception and reinterpretation of hypermetrical shifts rooted in phenomenology.

2. FORRÓ: STYLE AND HYPERMETERIC STRUCTURE

Forró can be regarded, along with many other hybrid forms, as an amalgamation between *baião* (as popularized in the Brazilian northeast by Luiz Gonzaga in the 1940s) and the Jamaican reggae. In its slow version, frequently referred to as *xote*, *forró* is a musical form that inherits its basic formative elements from European ballroom dances both in tonal and hypermetric structure. Thus, at a hypermetric level, it counts on duple, cyclic, metrical structural organizations. Generally, musical phrases will be multiples of 4 in number of bars, frequently displaying a [8 + 8] set up. The hypermetric structure thus, is perceived as falling on the downbeats of each set of 8 bars.

3. HYPERMETRIC SCTRUCTURE IN “ESPERANDO NA JANELA”

In the example analyzed in this abstract, however, the hypermeter suffers a shift, due to an elongation to the text and, consequently, to the stanza as a whole. Figure 1 illustrates the text and its musical accents. Figure 2 shows an organization of four metrical cells of four bars each, featuring a [(4 + 4 + 2) + (4 + 4)] organization. The symbols used in Figure 1 and 2 are the same so as to emphasize the metrical shift as occurring in text s well as in the overall structure.

The normal expectation is for m. 9 to be accented (Figure 2). Instead, the text has one more verse, which does not meet the listener’s “expectation” (as I will discuss below), thus elongating the second rhythmic cell. This results in a metrical shift delayed by two bars featuring, ultimately a [10 + 8] structure, hence disrupting the usual (or *expected*) metrical structure in *forró* music.

4. HYPERMETRIC SHIFTS AND PERCEPTION

Phenomenologically speaking, the perceptual experience of this significant metrical shift bears no immediate influence on dancers, as the dance steps of *forró* music are based on the lowest metrical unit: the beat. This hypermetric shift has implications, however, for the listener. Other examples are also found throughout other forms in tonal music repertoire of various places, throughout history. Since most examples of the *forró* form will have an [8 + 8] structure, this additional elongation is promptly noticeable and causes a building in tension as the refrain approaches.

The phenomenology of perception of hypermetric shifts, in cases like *forró* music—which normally relies on a very simple and uncomplicated metric structure—may be defined in terms of a three-step perceptual model that can be understood in time. This model is based on the culturally constructed economy of tonal musical (1) expectation, (2) reality, and (3) potential reinterpretation. Since the expectation is not met in this example, the careful listener is likely to feel the tension caused by the elongated metric cell. In facing the reality of the new hypermetric point, which now marks the beginning of the refrain, the structure then is reinterpreted *retrospectively*, phenomenologically speaking.

5. FINAL THOUGHTS

Hypermetric shifts can be recurrent in folkloric hybrid forms in any repertoire or style that derives from European forms. This is to say that deviations in culturally constructed examples of formal, metrical, and tonal structure can play a huge role in the perception of folkloric music in phenomenological terms.

6. REFERENCES

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

Temperley, D. (2008). Hypermetrical transitions. *Music Theory Spectrum, 20*(2), 305-325.

McCleland (2006). Extended upbeats in the classical minuet: interactions with hypermeter and phrase structure. *Music Theory Spectrum, 28*(1), 23-55.

“Esperando na Janela” (2000)

TEXT/POETRY

STANZA:

■
Ainda me lembro do seu cami**nhar**;
–
Seu jeito de olhar, eu me lembro **bem**
+
Fico querendo sentir o seu **chei**ro;
–
É daquele jeito que ela **tem**

■
O tempo todo eu fico feito **ton**to,
–
Sempre procurando, mas ela não **vem**
+
E esse aperto no fundo do **pei**to
–
Desses que o sujeito não pode aguen**tar**,(ah)

□
E esse aperto aumenta meu de**se**jo [e]
–
Eu não vejo a hora de poder lhe fa**lar**.

REFRAIN:

■
Por isso eu vou na casa **de**la, ai, ai
–
Falar do meu amor pra **e**la, vai;
+
Tá me esperando na ja**ne**la, ai, ai
–
Não sei se vou me segu**rar**.

■
Por isso eu vou na casa **de**la, ai, ai
–
Falar do meu amor pra **e**la, vai;
+
Tá me esperando na ja**ne**la, ai, ai
–
Não sei se vou me segu**rar**.

Figure 1. Poetry/text in “Esperando na Janela” (2000). Strong syllables are shown in bold and take place at the beginning of every musical bar.

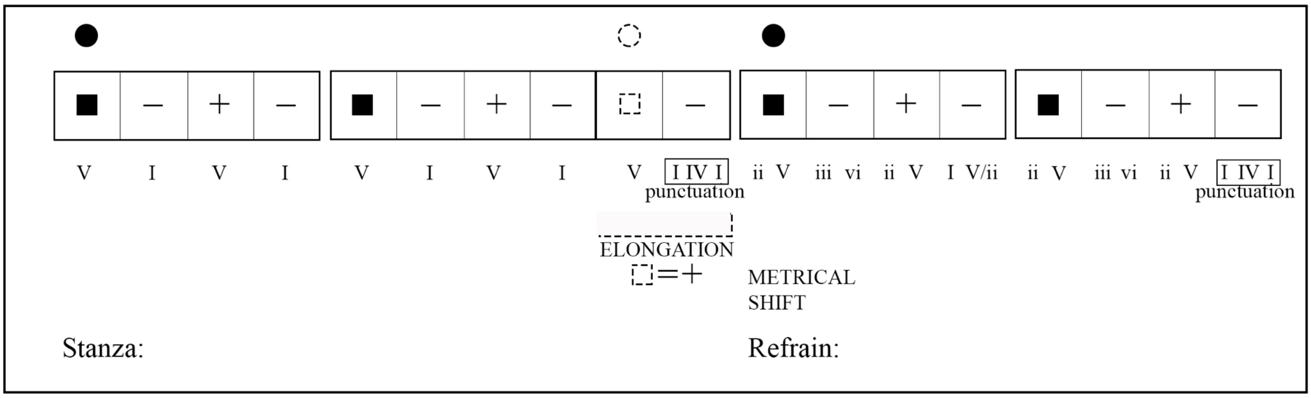


Figure 2. Hypermetric structure in “Esperando na Janela” (2000). Each individual block represents one (1) musical bar, normally organized in [(4 + 4) + (4 +4)] fashion, but here displaying an elongation. Black circles represent the hypermeter. Black squares indicate the accents at an intermediate level. The “+” and “-” symbols represent metrically strong and week bars, respectively, at an even lower metrical level.

RASPBERRY PI + LEGO = FOLK MUSIC DEMONSTRATOR

Alejandro Villena, Joaquín Cáceres, Marcelo Caetano, Isabel Barbancho, Lorenzo Tardón

ATIC Research Group, ETSI Telecomunicación
Universidad de Málaga (UMA), Andalucía Tech, Málaga, Spain
{avr,jcg,mcaetano,ibp,lorenzo}@ic.uma.es

ABSTRACT

Within the context of art music, it takes years of practice allied with formal musical education to master playing an instrument. Folk music, on the other hand, is transmitted from generation to generation by processes such as imitation and mimicking akin to oral tradition in storytelling [1]. In this work, we describe a modular system for music generation that allows tactile user interaction with musical parameters while providing both sonic and visual feedback. The system (generation component) uses 3D printed LEGO pieces to control MIDI [2] signals that propagate through the beat and melody generation modules synchronized by a master clock. The rhythm matrix provides visualization of the sonic output of the generation component in the style of a sequencer. User interaction combined with visualization leverages the power of practice in folk music to learn music theory intuitively. It is easy to generate musically interesting patterns and fun to interact with the system, which motivates the user to keep exploring the musical possibilities while learning.

Keywords: Interactive Music, MIDI, Modular Audio system, Musical education, Raspberry PI, LEGO

1. INTRODUCTION

It is widely agreed upon that mastering a musical instrument takes years of practice and musical education. On the other hand, folk music is traditionally transmitted orally from one generation to the next. Folk music commonly features relatively simple repeating patterns that lead to a complex musical structure when combined. There have been proposals that exploit Folk music's tradition to learn by practice. For example, the work described in [3] allows the user to load, play, and edit different rhythm loops separated by instruments.

The system proposed in this work uses hardware to manage the pattern-generating functionality. Raspberry PI 3 was chosen as the main brain for the system. Due to its '1.2GHz Quad-Core ARM Cortex-A53' processor, it can handle the MIDI protocol and hardware reading with no effort. It also provides flexibility for further functionalities such as buttons, LEDs or additional inputs/outputs. 3D printed LEGO bricks are the key to control said loops.

Figure 1 shows a block diagram illustrating the four core modules of the proposed system, namely the clock, the rhythm box, the melody box, and the VST. The clock generates the first data stream that controls all the subsequent elements. Both beatbox and melody box are able to repeat the input data to the output and adding its own data to the stream, which makes them independent of the position within the chain. The final element is a VST [4] that reads all the MIDI data and plays real sounds, according to the data received.

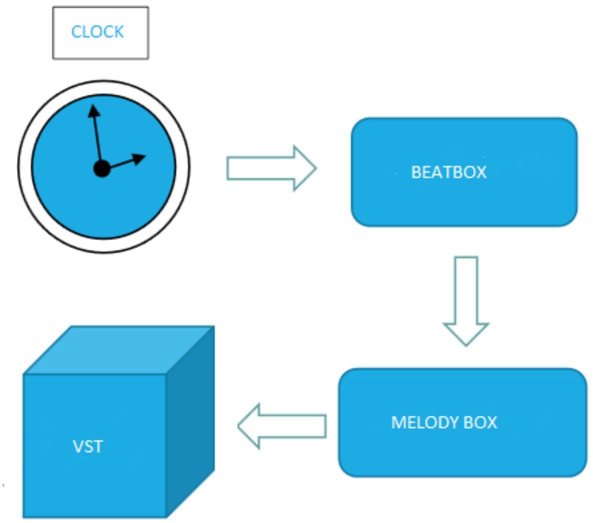


Figure 1: System chain description

2. SYSTEM DESCRIPTION

2.1 Clock

The clock controls the 'beats per minute' (BPM) of the loops. Its MIDI dictionary is not long, where the most important messages are 'timing clock' and 'start'. The 'timing clock' message is sent periodically. Every twenty-four of these messages, a quarter-note is played. That allows the beatbox and the melody box to play quarter notes at an exact BPM. The 'start' messages are sent at the beginning of every loop. The following elements wait for the first 'start' message to arrive, and use it to know when to start playing the pre-set pattern.

2.2 Beatbox

The next item in our chain is the beatbox. It repeats the MIDI stream coming from the clock. The beat loops are controlled by a 4 x 16 matrix similar to the one in Figure 2. Each row controls one type of beat. In the example, we could control the patterns of the kick, clap, closed hi-hat and open hi-hat. For each quarter-note, only one column of the matrix is read and played.

ISBN: 978-84-697-2303-6

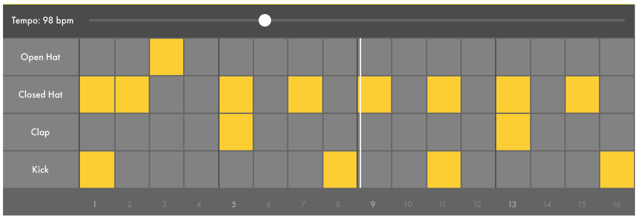


Figure 2: Rhythm matrix

By placing the LEGO bricks into one spot of the matrix, the spot gets "activated" and will generate the correspondent beat. Figure 3 features the same pattern as Figure 2, made out of LEGO bricks. One of the instruments can be sacrificed in order to use that row as an 'expression control' such as marking the strong beat within the loop, and making it sound louder.

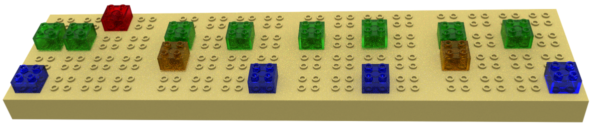


Figure 3: The Beatbox

2.3 Melody box

The melody box features the matrix described Figure 2, but the main difference is that the lower row can recognize stacked LEGO bricks as shown in Figure 4.

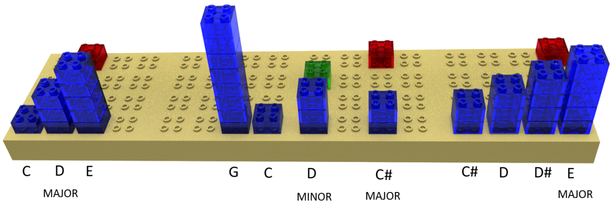


Figure 4: The melody box

Every stacked brick represents a half tone. Stacking LEGO pieces allows the user to intuitively create melodies using the visual interpretation that higher stacks generate higher pitches. Arpeggios and scales can be seen and spotted like a staircase. The first pentagram of Figure 5 illustrates the music transcription of just the lower row of the matrix. For the first four quarter-notes we have: (1st) one LEGO brick which means C, (2nd) three bricks which means D, (3rd) five bricks which means E and (4th) silence. The other rows control which chords are going to be generated. In our example, the 3rd row of the matrix generates minor chords when checked and the 4th generates major chords. Following the example, the second pentagram of Figure 5 now includes the newly generated major and minor chords.

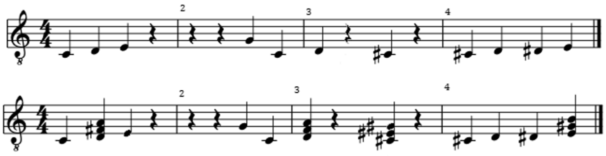


Figure 5: Music transcription

2.4 VST

The last part of the chain can be any VST. In our implementation, the VST chosen was a computer with a free DAW and VST plugins. The current implementation uses a drum machine for the beatbox and flute synth for melodies. The beatbox and the melody box are digitally split into 2 different MIDI channels. The drums are sent on channel 10 and melodies on channel 1, so that the whole data stream can be played by the same VST.

3. CONCLUSIONS

Currently, the system is mounted and running on the breadboard as shown in Figure 6. It is easy and fun to generate musical rhythms but it has certain limitations regarding what can be achieved musically. Ideally, the system should be able to generate more complex patterns and the user should be able to store a MIDI template. In the future, more tracks would be added to each module in order to increase functionality.

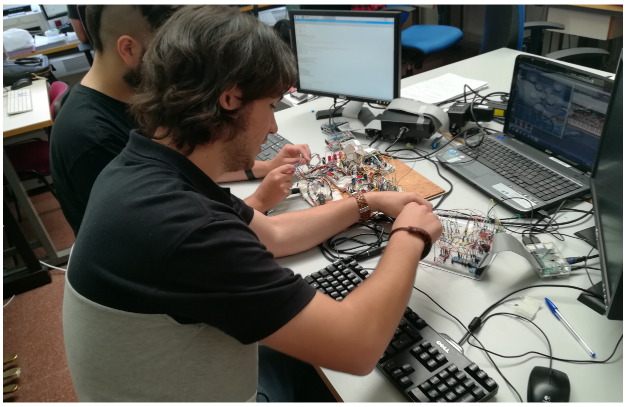


Figure 6: Breadboard prototype

4. REFERENCES

[1] Azahara Arévalo Galán. *Importance of Music folklore as educational practice*,Revista Electr. De LEEME, 23, 2009.

[2] The MIDI manufacturers association, Los Angeles CA (1996). V *The complete MIDI 1.0. Detailed Specification*

[3] Sevilla Soft (2012). The title of program is *MFFS_PROF*

[4] Luis Roberto Martinez Núñez (2007). *Procesamiento Digital y Control Gestural en tiempo real utilizando una PC con drivers ASIO para efectos de audio*. Chapter 6, dsp y vst plugin.

5. ACKNOWLEDGMENTS

This work has been funded by Ministerio de Economía y Competitividad of the Spanish Government under Project No. TIN2016-75866-C3-2-R. This work has been done at Universidad de Málaga, Campus de Excelencia Internacional Andalucía Tech.

Proceedings of the 7th International Workshop on Folk Music Analysis, 14-16 June 2017, Málaga (SPAIN)
ISBN: 978-84-697-2303-6