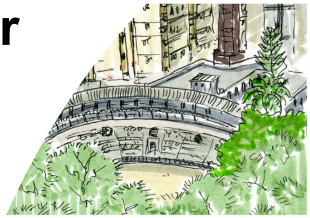
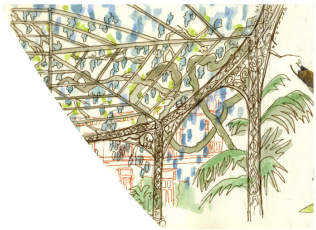
ISMIR 2015 MALAGA

Program & Abstracts

16th International Society for Music Information Retrieval Conference

October 26 - 30, 2015
Málaga, Spain





ISMIR 2015

Program and Abstracts of the
16th International Society for
Music Information Retrieval Conference

October 26 - 30, 2015
Málaga, Spain

ISMIR 2015 is organized by
International Society for Music Information Retrieval
ATIC Research Group. Universidad de Málaga

Website: http://ismir2015.ismir.net
http://ismir2015.uma.es

Program and Abstracts of the
16th International Society for Music Information Retrieval Conference
Cover design by Alberto Peinado & Isabel Barbancho
Cover Málaga drawings by Cristina Urdiales
Cover G-clef drawing by Alberto Peinado & Isabel Barbancho
ISMIR 2015 logo by Alberto Peinado & Emilio Molina

Edited by:
Isabel Barbancho (ATIC Research Group. Universidad de Málaga)
Lorenzo J. Tardón (ATIC Research Group. Universidad de Málaga)
Ana M. Barbancho (ATIC Research Group. Universidad de Málaga)

ISBN 978-84-608-2694-1

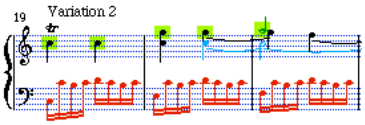
ORGANIZATION

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

UNIVERSIDAD DE MÁLAGA

iC Ingeniería de Comunicaciones

Escuela Técnica Superior de Ingeniería de Telecomunicación



ISMIR

SPONSORS

Gold Sponsors

Silver Sponsors

Bronze Sponsors

Other Collaborators

Contents

Welcome

Welcome to the 16th International Society for Music Information Retrieval Conference (ISMIR 2015) and Welcome to our nice city, Málaga. The annual ISMIR conference is the world's main research forum on MIR related topics. ISMIR is currently the 5th ranked publication in the "Multimedia" subcategory of "Engineering and Computer Science" and the 1st ranked in the "Music & Musicology" subcategory of "Humanities, Literature, and Arts".

ISMIR is not only a truly interdisciplinary conference, involving researchers from different disciplines, but also a really International Conference. This year we have over 225 participants from 25 different countries around the globe.

ISMIR 2015 conference has a most interesting TECHNICAL PROGRAM featuring:

- Six TUTORIALS that provide a good balance between culture and technology.
- Two KEYNOTE TALKS by two distinguished speakers: Prof. Mark Sandler & Prof. J. Stephen Downie.
- Seven ORAL SESSIONS.
- Three POSTER SESSIONS.
- WiMIR session aimed to promote the role of women in the MIR field.
- An INDUSTRIAL PANEL session that will reinforce the link between industry, research and academia.
- MIREX oral & poster sessions, in which state-of-the-art MIR methods are critically evaluated.
- BUSINESS MEETING where the ISMIR board members present their work and activities to the community since last ISMIR.

- LATE-BREAKING & DEMO (LBD) session, which will make all of us enjoy with brand new ideas and applications.

- UNCONFERENCE where ISMIR 2015 participants discuss MIR issues of particular interest in an informal context. This is a good opportunity to get to know peers and colleagues from all around the world.

ISMIR 2015 aims to offer to the participant an enriching experience not only from the technical point of view. The SOCIAL PROGRAM will provide participants with an opportunity to relax after meetings, to experience Málaga, and to network with other ISMIR participants:

- On Monday, October 26, we will get-together during the WELCOME RECEPTION at "Los Patios de Beatas".

- On Wednesday, October 28, at the "Sala UNICAJA de Conciertos María Cristina", we will enjoy a Live Flamenco Dancing & Music Show, we will hear music played by Sinfonietta of San Francisco de Paula and we will be surprised by the music from the "ISMIR 2015 Call for Music".

- On Thursday, October 29, we will celebrate ISMIR 2015 during the GALA DINNER at "Hacienda del Alamo". Awards will be given to outstanding ISMIR researchers. The grand finale of the dinner will be the ISMIR 2015 Jam Session. All ISMIR participants enjoy music, actually, many ISMIR attendees are either amateur or professional musicians. This year, the ISMIR participants will have, not only the opportunity to discuss novel MIR techniques but also the chance to play music together. Let's play some MUSIC!

- Also, during the conference, there will be room for relax, to interact with other ISMIR participants and even to play music between sessions. There will be a piano and a guitar for the ISMIR 2015 attendees who want to play (not very loud).

We want to thank the authors who sent their contributions, the reviewers, the meta-reviewers, the people attending the conference and the ISMIR community in general: you make this event possible.

We are grateful to the sponsors: FECYT (under project FCT-14-8217), Shazam, Gracenote, Pandora, Bose Corporation, Universidad de Málaga, MINECO (under project TIN2015-62946-CIN), Steinberg Media Technologies GmbH, Native Instruments GmbH, Google, ACRCloud, Smule, Departamento

Ingeniería de Comunicaciones, E.T.S.I.Telecomunicación, Fundación Unicaja, Ayuntamiento de Málaga and Málaga Convention Bureau for their support.

Special thanks go to all the members of the ISMIR 2015 Conference Committee: Program Chairs, Tutorial Chair, Unconference Chairs, Late-Breaking and Demo Chairs, Music Chairs, Music Curator, Jam Session Chairs, Local Arrangement Chairs and Program Committee. Our most sincere acknowledgment is for the Program Chairs Frans Wiering & Meinard Müller because their work has been of key importance for the success of the ISMIR 2015 conference. Last, but no least, we want to specially thank the Local Arrangement Chairs Alberto Peinado, for his help with the graphical design of all the ISMIR 2015 posters, and Ana M. Barbancho, for her help in absolutely everything.

It is a great honor to be in charge of the organization of the ISMIR 2015 conference and, at the same time, it has been a big challenge. We hope that our effort will be worthy and that we all enjoy a great conference. Now, ISMIR 2015 is in your hands. Let's do the most of it. Let's make ISMIR 2015 an unforgettable experience!

Málaga,
October 2015

Isabel Barbancho
Lorenzo J. Tardón
General Chairs, ISMIR 2015

Preface

The annual ISMIR conference is the world's leading research forum on processing, analyzing, searching, organizing, and accessing music-related data. This year's conference, which takes place in Málaga, Spain, October 26-30, 2015, is organized by the ATIC Research Group, Universidad de Málaga. The present volume contains the conference program and the abstracts of all the peer-reviewed papers to be presented at ISMIR 2015. A total of 278 submissions were received before the deadline, out of which 242 complete and well-formatted papers entered the review process. Special care was taken to assemble an experienced and interdisciplinary review panel including people from many different institutions worldwide. As in previous years, the reviews were double-blinded (i.e., both the authors and the reviewers were anonymous) with a two-tier review model involving a pool of 257 reviewers and a program committee. Reviewers and PC members were able to bid for papers. Each paper was assigned to a PC member and three reviewers. Reviewer assignments were based on topic preferences, bids, and PC member assignments. After the review phase, PC members and reviewers entered a (name-disclosed) discussion phase aiming to homogenize acceptance vs. rejection decisions. Compared to previous years, the size of the program committee increased significantly and now comprises 61 members. Taking care of four submissions on average, the PC members were asked to adopt an active role in the review process by conducting an intensive discussion phase with the other reviewers and by providing a detailed meta-review. Final acceptance decisions were based on 973 reviews and meta-reviews. From the 242 reviewed papers, 114 papers were accepted resulting in an acceptance rate of 47.1%.

Year	Location	Oral	Poster	Total Papers	Total Pages	Total Authors	Unique Authors	Pages/ Paper	Authors/ Paper	Unique Authors/ Paper
2000	Plymouth	19	16	35	155	68	63	4.4	1.9	1.8
2001	Indiana	25	16	41	222	100	86	5.4	2.4	2.1
2002	Paris	35	22	57	300	129	117	5.3	2.3	2.1
2003	Baltimore	26	24	50	209	132	111	4.2	2.6	2.2
2004	Barcelona	61	44	105	582	252	214	5.5	2.4	2
2005	London	57	57	114	697	316	233	6.1	2.8	2
2006	Victoria	59	36	95	397	246	198	4.2	2.6	2.1
2007	Vienna	62	65	127	486	361	267	3.8	2.8	2.1
2008	Philadelphia	24	105	105	630	296	253	6	2.8	2.4
2009	Kobe	38	85	123	729	375	292	5.9	3	2.4
2010	Utrecht	24	86	110	656	314	263	6	2	2.4
2011	Miami	36	97	133	792	395	322	6	3	2.4
2012	Porto	36	65	101	606	324	264	6	3.2	2.6
2013	Curitiba	31	67	98	587	395	236	5.9	3	2.4
2014	Taipei	33	73	106	635	343	271	6	3.2	2.6
2015	**Málaga**	**24**	**90**	**114**	**792**	**370**	**296**	**7**	**3.2**	**2.6**

The table summarizes the ISMIR publication statistics over the last years. The mode of presentation of the papers was determined after the accept/reject decisions and has no relation to the quality of the papers or to the number of pages allotted in the proceedings. From the 114 contributions, 24 papers were chosen for oral presentation based on the topic and broad appeal of the work, whereas the other 90 were chosen for poster presentation. Oral presentations have a 20-minute slot (including setup and questions/answers from the audience) whereas poster presentations are done in two sessions per day, the same posters being presented in the morning and in the afternoon of a given conference day.

In this book, you will find the detailed program of ISMIR 2015, together with the abstracts of all the contributions that will be presented during the oral and poster sessions and information about all the technical activities and other sessions that will take place during the conference days. The information on the Late-Breaking/Demo session will be available at the ISMIR 2015 website (http://ismir2015.uma.es).

Enjoy ISMIR 2015!

Málaga,
October 2015

Isabel Barbancho
Lorenzo J. Tardón
General Chairs, ISMIR 2015

Meinard Müller
Frans Wiering
Program Chairs, ISMIR 2015

Conference Committee

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

Program Chairs
Meinard Müller (International Audio Laboratories Erlangen, Germany)
Frans Wiering (Utrecht University, The Netherlands)

Tutorial Chair
George Tzanetakis (University of Victoria, Canada)

Unconference Chairs
Dan Ellis (Columbia University, USA)
Douglas Eck (Google)

Late-Breaking and Demo Chairs
Christian Dittmar (International Audio Laboratories Erlangen, Germany)
Ana M. Barbancho (Universidad de Málaga, Spain)

Music Chairs
José Pedro Rodrigues Magalhães (Chordify, The Netherlands)
Carlos Guedes (New York University, Abu Dhabi)

Music Curator
Robert Rowe (New York University, USA)

Jam Session Chairs
Oriol Nieto (Pandora, USA)
Emilio Molina (Universidad de Málaga, Spain)

Local Arrangement Chairs
Alberto Peinado (Universidad de Málaga, Spain)
Andrés Ortiz (Universidad de Málaga, Spain)
Jorge Munilla (Universidad de Málaga, Spain)
José L. Santacruz (Universidad de Málaga, Spain)
Jesús Corral (Universidad de Málaga, Spain)
Hande Başak (Anadolu University, Turkey)
José Manuel Peula Palacios (Universidad de Málaga, Spain)
Manuel Fernández Carmona (Universidad de Málaga, Spain)
Joaquín Ballesteros (Universidad de Málaga, Spain)

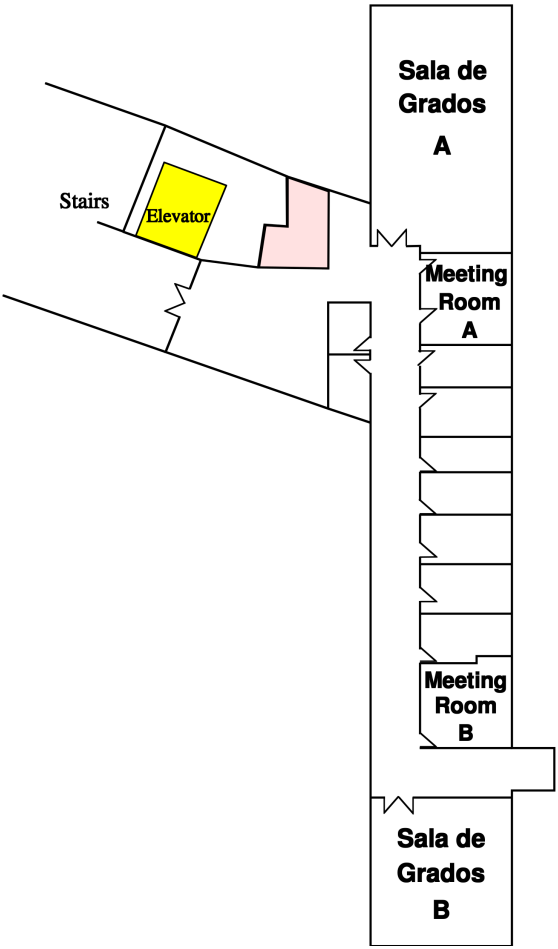
Program Committee

Amélie Anglade, Freelance MIR and Data Science Consultant, Germany
Jean-Julien Aucouturier, CNRS/IRCAM, France
Eric Battenberg, Gracenote, USA
Juan Pablo Bello, New York University, USA
Emmanouil Benetos, Queen Mary University of London, UK
Dawn Black, Queen Mary University of London, UK
Sebastian Böck, Johannes Kepler University, Austria
Ashley Burgoyne, University of Amsterdam, The Netherlands
Ching-Hua Chuan, University of North Florida, USA
Tom Collins, De Montfort University, UK
Sally Jo Cunningham, The University of Waikato, New Zealand
Roger Dannenberg, Carnegie Mellon University, USA
Matthew Davies, INESC TEC, Portugal
Johanna Devaney, The Ohio State University, USA
Christian Dittmar, AudioLabs, Germany
Simon Dixon, Queen Mary University of London, UK
J. Stephen Downie, University of Illinois at Urbana-Champaign, USA
Zhiyao Duan, University of Rochester, USA
Andreas Ehmann, Pandora, USA
Dan Ellis, Columbia University, USA
Sebastian Ewert, Queen Mary University of London, UK
George Fazekas, Queen Mary University of London, UK
Arthur Flexer, Austrian Research Institute for Artificial Intel. (OFAI), Austria
José Fornari, NICS - UNICAMP, Brasil
Ichiro Fujinaga, McGill University, Canada
Emilia Gómez, Universitat Pompeau Fabra, Spain
Masataka Goto, AIST, Japan
Fabien Gouyon, Pandora, USA
Maarten Grachten, OFAI, Austria
Bas de Haas, Utrecht University, The Netherlands

Dorien Herremans, Universiteit Antwerpen, Belgium
Perfecto Herrera, Universitat Pompeu Fabra, Spain
Andre Holzapfel, Bogazici University, Turkey
Xiao Hu, University of Hong Kong, Hong Kong
Eric Humphrey, MuseAmi/NYU, USA
Ozgur Izmirli, Connecticut College, USA
Peter Knees, Johannes Kepler University Linz, Austria
Peter van Kranenburg, Meertens Institute, Amsterdam, The Netherlands
Audrey Laplante, Université de Montréal, Canada
Jinha Lee, University of Washington, USA
Cynthia Liem, Delft University of Technology, The Netherlands
Matthias Mauch, Queen Mary University of London, UK
Matt McVicar, The University of Bristol, UK
Nicola Orio, University of Padua, Italy
Geoffroy Peeters, UMR IRCAM CNRS STMS, France
Marcelo Queiroz, University of Sao Paulo, Brasil
Christophe Rhodes, Goldsmiths, University of London, UK
Justin Salamon, New York University, USA
Erik Schmidt, Pandora, USA
Jeffrey Scott, Gracenote, USA
Xavier Serra, Universitat Pompeu Fabra, Spain
Jordan Smith, AIST, Japan
Mohamed Sordo, University of Miami, USA
Bob Sturm, Aalborg University, Denmark
Li Su, Academia Sinica, Taiwan
Douglas Turnbull, Ithaca College, USA
George Tzanetakis, University of Victoria, Canada
Julian Urbano, Universitat Pompeu Fabra, Spain
Ju-Chiang Wang, Academia Sinica, Taiwan
Yi-Hsuan Yang, Academia Sinica, Taiwan

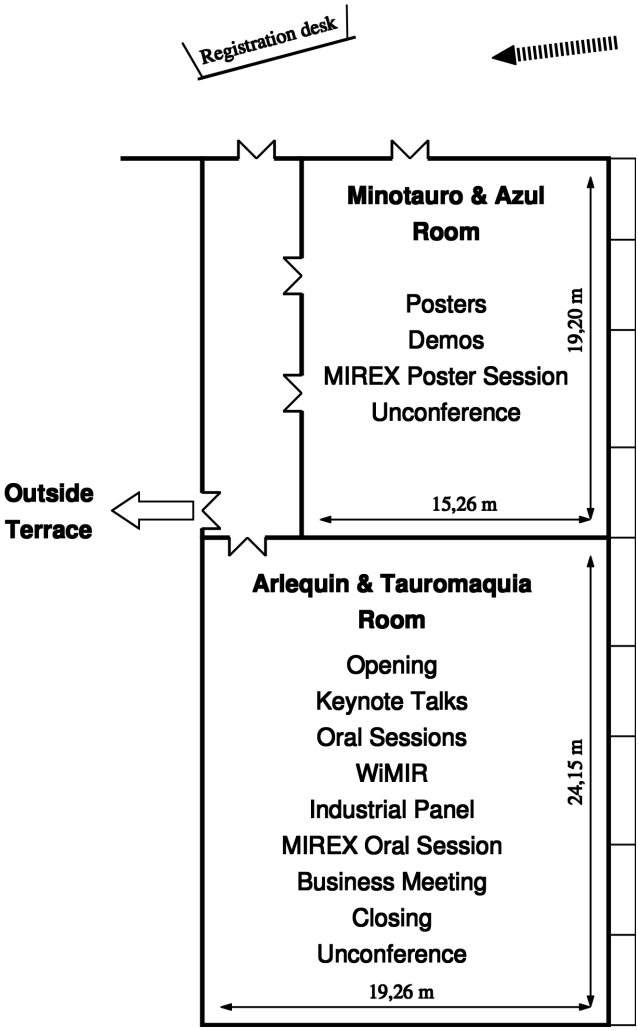
Conference Venue

ISMIR 2015 Tutorial Venue
E.T.S.I. Telecomunicación
(*http://www.etsit.uma.es*)
Campus de Teatinos s/n, 29071 Málaga, Spain
Tel. +34 952 132 700



Tutorial floor map.
E.T.S.I. Telecomunicación, fourth floor.

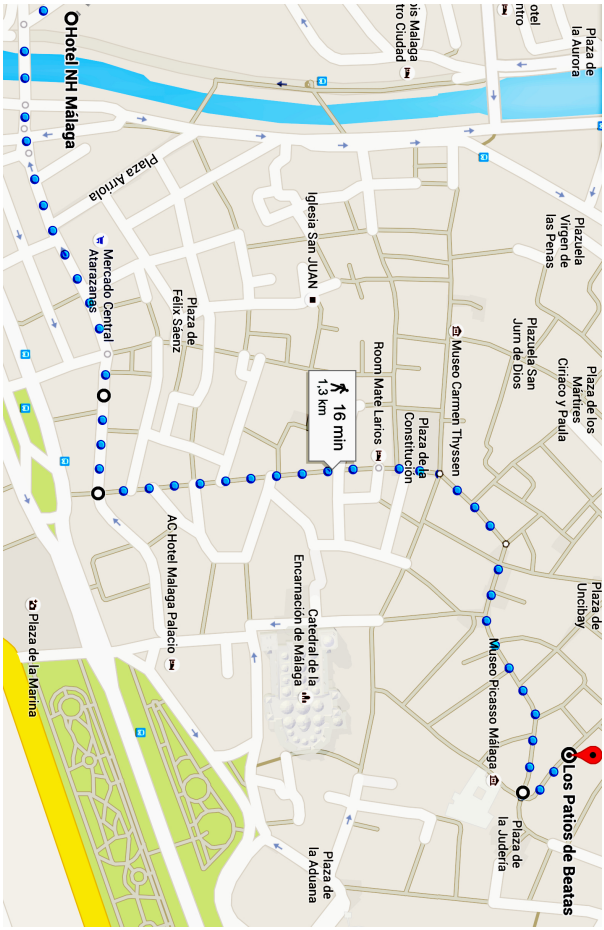
ISMIR 2015 Conference Venue
Hotel NH Málaga
(*http://www.nh-hoteles.es/hotel/nh-malaga*)
Calle San Jacinto 2, 29007 Málaga, Spain
Tel. +34 952 071 323



Conference floor map.
Hotel NH Málaga, lower floor.

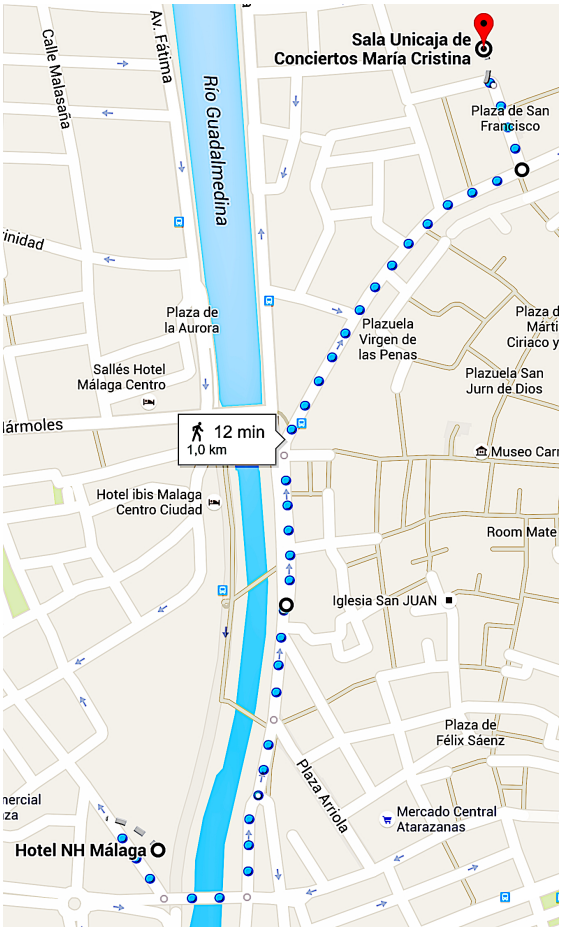
Social Program Venues

ISMIR 2015 Welcome Reception
Vinoteca Museo Los Patios de Beatas
(*http://lospatiosdebeatas.com/*)
Calle Beatas, 43, 29006 Málaga, Spain
Tel. +34 952 210 350



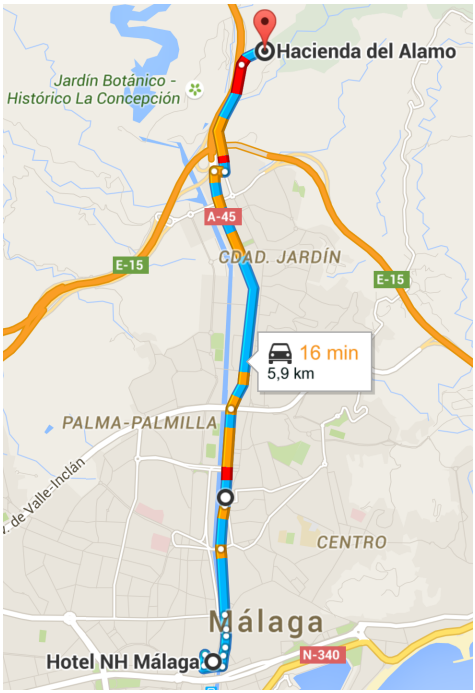
Route from
Conference Venue-Hotel NH Malaga
to
Vinoteca Museo Los Patios de Beatas.

ISMIR 2015 Flamenco Show & Concert
Sala Unicaja de Conciertos María Cristina
(*https://www.obrasocialunicaja.es/cultura/nuestros-centros/sala-unicaja-de-conciertos-maria-cristina/*)
Calle Marqués de Valdecañas 2, 29008 Málaga, Spain
Tel. +34 952 219 533



Route from
Conference Venue-Hotel NH Malaga
to
Sala Unicaja de Conciertos María Cristina.

ISMIR 2015 Gala Dinner & Pandora Jam Session
Hacienda del Alamo
(*https://www.haciendadelalamo.es/*)
Camino de Casabermeja, 130, 29014 Málaga, Spain
Tel. +34 952 652 653



Route from
Conference Venue-Hotel NH Malaga
to
Hacienda del Alamo.

(Bus transportation will be provided at 19:30 from *Hotel NH Málaga* to *"Hacienda del Alamo"* and back to *Hotel NH Málaga* after the ISMIR 2015 Gala Dinner & Pandora Jam Session)

Conference Schedule

Program at a Glance

Monday, October 26 Venue: E.T.S.I. Telecomunicación

09:00-10:00	Registration
10:00-11:30	**Tutorials 1, 2 and 3**
11:30-12:00	Coffee
12:00-13:30	**Tutorials 1, 2 and 3**
13:30-14:30	Break
14:30-16:00	**Tutorials 4, 5 and 6**
16:00-16:30	Coffee
16:30-18:00	**Tutorials 4, 5 and 6**
20:00	Welcome Reception

Tuesday, October 27 Venue: Hotel NH Málaga

09:00-10:00	Registration
10:00-10:30	**Opening**
10:30-11:30	**Keynote Speaker 1**
11:30-13:30	**Poster Session 1** & Coffee
13:30-14:30	Lunch
14:30-16:00	**Oral Session 1. Corpus Analysis & Annotation**
16:00-17:30	**Poster Session 1** & Coffee

Wednesday, October 28 Venue: Hotel NH Málaga

09:00-10:00	**Oral Session 2. Rhythm & Beat**
10:00-11:30	**Poster Session 2** & Coffee
11:30-12:30	**Oral Session 3. Melody & Voice**
12:30-13:30	**WiMIR**
13:30-14:30	Lunch
14:30-16:00	**Oral Session 4. Mixed**
16:00-17:30	**Poster Session 2** & Coffee
18:00	Flamenco Show & Concert

Thursday, October 29 Venue: Hotel NH Málaga

09:00-10:00	**Oral Session 5. Similarity**
10:00-11:30	**Poster Session 3** & Coffee
11:30-12:30	**Keynote Speaker 2**
12:30-13:30	**Industrial Panel**
13:30-14:30	Lunch
14:30-16:00	**Oral Session 6. User & Communnity**
16:00-17:30	**Poster Session 3** & Coffee
20:00	ISMIR 2015 Gala Dinner & Pandora Jam Session

Friday, October 30 Venue: Hotel NH Málaga

09:00-10:00	**Oral Session 7. Performance**
10:00-11:00	**MIREX Oral Session**
11:00-12:30	**MIREX Poster Session** & Coffee
12:30-13:30	**Business Meeting**
13:30-13:40	**Closing**
13:40-14:30	Lunch
14:30-16:00	**Demos & Late-Breaking News**
16:00-17:30	**Unconference**

Keynote Speakers

We are honored to have two distinguished keynote speakers:

- Prof. Mark Sandler from Queen Mary University of London, UK.
- Prof. J. Stephen Downie from the University of Illinois at Urbana-Champaign, USA.

Tutorials

Monday, October 26 Venue: E.T.S.I. Telecomunicación

09:00-10:00	Registration
10:00-11:30	**Tutorials 1, 2 and 3**

Tutorial 1 **Why Singing is Interesting**
Simon Dixon, Masataka Goto, Matthias Mauch
Location: Sala de Grados A

Tutorial 2 **Addressing the Music Information Needs of Musicologists**
Richard J. Lewis, Ben Fields, Tim Crawford
Location: Sala de Grados B

Tutorial 3 **Markov Logic Networks for Music Analysis**
Helene Papadopoulos
Location: Meeting Room B

11:30-12:00	Coffee
12:00-13:30	**Tutorials 1, 2 and 3**
13:30-14:30	Break
14:30-16:00	**Tutorials 4, 5 and 6**

Tutorial 4 **COmputation and FLAmenco: Why Flamenco is Interesting for MIR Research**
Emilia Gómez, Nadine Kroher, José Miguel Díaz-Báñez, Sergio Oramas, Joaquín Mora, Francisco Gómez-Martín
Location: Meeting Room B

Tutorial 5 **Using Correlation Analysis and Big Data to Identify and Predict Musical Behaviors**
Jeff C. Smith
Location: Sala de Grados B

Tutorial 6 **Automatic Music Transcription**
Zhiyao Duan, Emmanouil Benetos
Location: Sala de Grados A

16:00-16:30	Coffee
16:30-18:00	**Tutorials 4, 5 and 6**
20:00	Welcome Reception

Conference Program

Tuesday, October 27 Venue: Hotel NH Málaga

09:00-10:00	Registration
10:00-10:30	**Opening**
10:30-11:30	**Keynote Speaker 1**
	Integrating Music Information Sources for Music Production and Consumption *Mark Sandler*
11:30-13:30	**Poster Session 1** & Coffee

PS1.1 Image Quality Estimation for Multi-Score OMR
Dan Ringwalt, Roger B. Dannenberg

PS1.2 Comparative Music Similarity Modelling Using Transfer Learning Across User Groups
Daniel Wolff, Andrew MacFarlane, Tillman Weyde

PS1.3 Modeling Genre with the Music Genome Project: Comparing Human-Labeled Attributes and Audio Features
Matthew Prockup, Andreas F. Ehmann, Fabien Gouyon, Erik M. Schmidt, Oscar Celma, Youngmoo E. Kim

PS1.4 Cover Song Identification with Timbral Shape Sequences
Christopher J. Tralie, Paul Bendich

PS1.5 On the Impact of Key Detection Performance for Identifying Classical Music Styles
Christof Weiß, Maximilian Schaab

PS1.6 Chord Detection Using Deep Learning
Xinquan Zhou, Alexander Lerch

PS1.7 Temporal Music Context Identification with User Listening Data
Cameron Summers, Phillip Popp

PS1.8 Improving Music Recommendations with a Weighted Factorization of the Tagging Activity
Andreu Vall, Marcin Skowron,Peter Knees, Markus Schedl

PS1.9 An Efficient State-Space Model for Joint Tempo and Meter Tracking
Florian Krebs, Sebastian Böck, Gerhard Widmer

PS1.10 Automatic Handwritten Mensural Notation Interpreter: From Manuscript to MIDI Performance
Yu-Hui Huang, Xuanli Chen, Serafina Beck, David Burn, Luc Van Gool

PS1.11 Infinite Superimposed Discrete All-Pole Modeling for Multipitch Analysis of Wavelet Spectrograms
Kazuyoshi Yoshii, Katsutoshi Itoyama, Masataka Goto

PS1.12 Melodic Similarity in Traditional French-Canadian Instrumental Dance Tunes
Laura Risk, Lillio Mok, Andrew Hankinson, Julie Cumming

PS1.13 A Semantic-Based Approach for Artist Similarity
Sergio Oramas, Mohamed Sordo, Luis Espinosa-Anke, Xavier Serra

PS1.14 Predicting Pairwise Pitch Contour Relations Based on Linguistic Tone Information in Beijing Opera Singing
Shuo Zhang, Rafael Caro Repetto, Xavier Serra

PS1.15 Song2Quartet: A System for Generating String Quartet Cover Songs from Polyphonic Audio of Popular Music
Graham Percival, Satoru Fukayama, Masataka Goto

PS1.16 Exploring Data Augmentation for Improved Singing Voice Detection with Neural Networks
Jan Schlüter, Thomas Grill

PS1.17 Audio Chord Recognition with a Hybrid Recurrent Neural Network
Siddharth Sigtia, Nicolas Boulanger-Lewandowski, Simon Dixon

PS1.18 Design and Evaluation of a Probabilistic Music Projection Interface
Beatrix Vad, Daniel Boland, John Williamson, Roderick Murray-Smith, Peter Berg Steffensen

PS1.19 Conceptual Blending in Music Cadences: A Formal Model and Subjective Evaluation
Asterios Zacharakis, Maximos Kaliakatsos-Papakostas, Emilios Cambouropoulos

PS1.20 Harmonic-Percussive Source Separation Using Harmonicity and Sparsity Constraints
Jeongsoo Park, Kyogu Lee

PS1.21 A Hierarchical Bayesian Framework for Score-Informed Source Separation of Piano Music Signals
Wai Man Szeto, Kin Hong Wong

PS1.22 Automatic Tune Family Identification by Musical Sequence Alignment
Patrick E. Savage, Quentin D. Atkinson

PS1.23 Evaluation of Album Effect for Feature Selection in Music Genre Recognition
Igor Vatolkin, Günter Rudolph, Claus Weihs

PS1.24 Music Pattern Discovery with Variable Markov Oracle: A Unified Approach to Symbolic and Audio Representations
Cheng-i Wang, Jennifer Hsu, Shlomo Dubnov

PS1.25 Automatic Solfège Assessment
Rodrigo Schramm, Helena de Souza Nunes, Cláudio Rosito Jung

PS1.26 Evaluating Conflict Management Mechanisms for Online Social Jukeboxes
Felipe Vieira, Nazareno Andrade

PS1.27 Particle Filters for Efficient Meter Tracking with Dynamic Bayesian Networks
Ajay Srinivasamurthy, Andre Holzapfel, Ali Taylan Cemgil, Xavier Serra

PS1.28 Predictive Power of Personality on Music-Genre Exclusivity
Jotthi Bansal, Matthew Woolhouse

PS1.29 A Toolkit for Live Annotation of Opera Performance: Experiences Capturing Wagner's *Ring* Cycle
Kevin R. Page, Terhi Nurmikko-Fuller, Carolin Rindfleisch, David M. Weigl, Richard J. Lewis, Laurence Dreyfus, David De Roure

PS1.30 Selective Acquisition Techniques for Enculturation-Based Melodic Phrase Segmentation
Marcelo E. Rodríguez-López, Anja Volk

13:30-14:30 Lunch

14:30-16:00 **Oral Session 1. Corpus Analysis & Annotation**
Session Chair: Cynthia C.S. Liem

OS1.1 Corpus Analysis Tools for Computational Hook Discovery
Jan Van Balen, John Ashley Burgoyne, Dimitrios Bountouridis, Daniel Müllensiefen, Remco C. Veltkamp

OS1.2 Large-Scale Content-Based Matching of MIDI and Audio Files
Colin Raffel, Daniel P. W. Ellis

OS1.3 Improving Genre Annotations for the Million Song Dataset
Hendrik Schreibers

OS1.4 A Software Framework for Musical Data Augmentation
Brian McFee, Eric J. Humphrey, Juan P. Bello

16:00-17:30 **Poster Session 1** & Coffee

Wednesday, October 28

Venue: Hotel NH Málaga

09:00-10:00 **Oral Session 2. Rhythm & Beat**
Session Chair: Sebastian Böck

OS2.1 Drum Transcription Using Partially Fixed Non-Negative Matrix Factorization with Template Adaptation
Chih-Wei Wu, Alexander Lerch

OS2.2 Beat and Downbeat Tracking Based on Rhythmic Patterns Applied to the Uruguayan Candombe Drumming
Leonardo Nunes, Martín Rocamora, Luis Jure, Luiz W. P. Biscainho

OS2.3 Automated Estimation of Ride Cymbal Swing Ratios in Jazz Recordings
Christian Dittmar, Martin Pfleiderer, Meinard Müller

10:00-11:30 **Poster Session 2** & Coffee

PS2.1 Musical Offset Detection of Pitched Instruments: The Case of Violin
Che-Yuan Liang, Li Su, Yi-Hsuan Yang, Hsin-Ming Lin

PS2.2 Specter: Combining Music Information Retrieval with Sound Spatialization
Bill Manaris, Seth Stoudenmier

PS2.3 Content-Aware Collaborative Music Recommendation Using Pre-trained Neural Networks
Dawen Liang, Minshu Zhan, Daniel P. W. Ellis

PS2.4 Comparative Analysis of Orchestral Performance Recordings: An Image-Based Approach
Cynthia C. S. Liem, Alan Hanjalic

PS2.5 Monaural Blind Source Separation in the Context of Vocal Detection
Bernhard Lehner, Gerhard Widmer

PS2.6 Detection of Common Mistakes in Novice Violin Playing
Yin-Jyun Luo, Li Su, Yi-Hsuan Yang, Tai-Shih Chi

PS2.7 Probabilistic Modular Bass Voice Leading in Melodic Harmonisation
Dimos Makris, Maximos Kaliakatsos-Papakostas, Emilios Cambouropoulos

PS2.8 An Iterative Multi Range Non-Negative Matrix Factorization Algorithm for Polyphonic Music Transcription
Anis Khlif, Vidhyasaharan Sethu

PS2.9 Training Phoneme Models for Singing with "Songified" Speech Data
Anna M. Kruspe

PS2.10 Graph-Based Rhythm Interpretation
Rong Jin, Christopher Raphael

PS2.11 Let it Bee – Towards NMF-Inspired Audio Mosaicing
Jonathan Driedger, Thomas Prätzlich, Meinard Müller

PS2.12 Real-Time Music Tracking Using Multiple Performances as a Reference
Andreas Arzt, Gerhard Widmer

PS2.13 Two Data Sets for Tempo Estimation and Key Detection in Electronic Dance Music Annotated from User Corrections
Peter Knees, Ángel Faraldo, Perfecto Herrera, Richard Vogl, Sebastian Böck, Florian Hörschläger, Mickael Le Goff

PS2.14 Towards Support for Understanding Classical Music: Alignment of Content Descriptions on the Web
Taku Kuribayashi, Yasuhito Asano, Masatoshi Yoshikawa

PS2.15 FlaBase: Towards the Creation of a Flamenco Music Knowledge Base
Sergio Oramas, Francisco Gómez, Emilia Gómez, Joaquín Mora

PS2.16 Discovery of Syllabic Percussion Patterns in Tabla Solo Recordings
Swapnil Gupta, Ajay Srinivasamurthy, Manoj Kumar, Hema A. Murthy, Xavier Serra

PS2.17 Autoregressive Hidden Semi-Markov Model of Symbolic Music Performance for Score Following
Eita Nakamura, Philippe Cuvillier, Arshia Cont, Nobutaka Ono, Shigeki Sagayama

PS2.18 Automatic Mashup Creation by Considering both Vertical and Horizontal Mashabilities
Chuan-Lung Lee, Yin-Tzu Lin, Zun-Ren Yao, Feng-Yi Lee, Ja-Ling Wu

PS2.19 Hierarchical Evaluation of Segment Boundary Detection
Brian McFee, Oriol Nieto, Juan P. Bello

PS2.20 Improving MIDI Guitar's Accuracy with NMF and Neural Net
Masaki Otsuka, Tetsuro Kitahara

PS2.21 Analysis of Intonation Trajectories in Solo Singing
Jiajie Dai, Matthias Mauch, Simon Dixon

PS2.22 Evaluating the General Chord Type Representation in Tonal Music and Organising GCT Chord Labels in Functional Chord Categories
Maximos Kaliakatsos-Papakostas, Asterios Zacharakis, Costas Tsougras,
Emilios Cambouropoulos

PS2.23 Beat Histogram Features from NMF-Based Novelty Functions for Music Classification
Athanasios Lykartsis, Chih-Wei Wu, Alexander Lerch

PS2.24 Music Shapelets for Fast Cover Song Recognition
Diego F. Silva, Vinícius M. A. Souza, Gustavo E. A. P. A. Batista

PS2.25 Improving Score-Informed Source Separation for Classical Music through Note Refinement
Marius Miron, Julio José Carabias-Orti, Jordi Janer

PS2.26 In their Own Words: Using Text Analysis to Identify Musicologists' Attitudes towards Technology
Charles Inskip, Frans Wiering

PS2.27 Combining Features for Cover Song Identification
Julien Osmalskyj, Peter Foster, Simon Dixon, Jean-Jacques Embrechts

PS2.28 Score Following for Piano Performances with Sustain-Pedal Effects
Bochen Li, Zhiyao Duan

PS2.29 Understanding Users of Commercial Music Services through Personas: Design Implications
Jin Ha Lee, Rachel Price

PS2.30 Corpus-Based Rhythmic Pattern Analysis of Ragtime Syncopation
Hendrik Vincent Koops, Anja Volk, W. Bas de Haas

11:30-12:30 **Oral Session 3. Melody & Voice**
Session Chair: Matthias Mauch

OS3.1 Comparing Voice and Stream Segmentation Algorithms
Nicolas Guiomard-Kagan, Mathieu Giraud, Richard Groult, Florence Levé

OS3.2 Melody Extraction by Contour Classification
Rachel M. Bittner, Justin Salamon, Slim Essid, Juan P. Bello

OS3.3 Comparison of the Singing Style of Two Jingju Schools
Rafael Caro Repetto, Rong Gong, Nadine Kroher, Xavier Serra

12:30-13:30	**WiMIR** *Hosts: Amélie Anglade, Emilia Gómez, Jin Ha Lee, Anja Volk*
13:30-14:30	Lunch
14:30-16:00	**Oral Session 4. Mixed** *Session Chair: Eric J. Humphrey*

OS4.1 Improving Optical Music Recognition by Combining Outputs from Multiple Sources
Victor Padilla, Alex McLean, Alan Marsden, Kia Ng

OS4.2 Relating Natural Language Text to Musical Passages
Richard Sutcliffe, Tim Crawford, Chris Fox, Deane L. Root, Eduard Hovy, Richard J. Lewis

OS4.3 Music Boundary Detection Using Neural Networks on Combined Features and Two-Level Annotations
Thomas Grill, Jan Schlüter

OS4.4 Neuroimaging Methods for Music Information Retrieval: Current Findings and Future Prospects
Blair Kaneshiro, Jacek P. Dmochowski

16:00-17:30	**Poster Session 2** & Coffee
18:00	Flamenco Show & Concert

Thursday, October 29

Venue: Hotel NH Málaga

09:00-10:00	**Oral Session 5. Similarity** *Session Chair: Bob Sturm*

OS5.1 Improving Visualization of High-Dimensional Music Similarity Spaces
Arthur Flexer

OS5.2 I-Vectors for Timbre-Based Music Similarity and Music Artist Classification
Hamid Eghbal-zadeh, Bernhard Lehner, Markus Schedl, Gerhard Widmer

OS5.3 Correlating Extracted and Ground-Truth Harmonic Data in Music Retrieval Tasks
Dylan Freedman, Eddie Kohler, Hans Tutschku

10:00-11:30 **Poster Session 3** & Coffee

PS3.1 Classical Music on the Web – User Interfaces and Data Representations
Martin Gasser, Andreas Arzt, Thassilo Gadermaier, Maarten Grachten, Gerhard Widmer

PS3.2 A Statistical View on the Expressive Timing of Piano Rolled Chords
Mutian Fu, Guangyu Xia, Roger B. Dannenberg, Larry Wasserman

PS3.3 Hybrid Long- and Short-Term Models of Folk Melodies
Srikanth Cherla, Son N. Tran, Tillman Weyde, Artur d'Avila Garcez

PS3.4 Efficient Melodic Query Based Audio Search for Hindustani Vocal Compositions
Kaustuv Kanti Ganguli, Abhinav Rastog, Vedhas Pandit, Prithvi Kantan, Preeti Rao

PS3.5 Modified Perceptual Linear Prediction Liftered Cepstrum (MPLPLC) Model for Pop Cover Song Recognition
Ning Chen, J. Stephen Downie, Haidong Xiao, Yu Zhu, Jie Zhu

PS3.6 Raga Verification in Carnatic Music Using Longest Common Segment Set
Shrey Dutta, Krishnaraj Sekhar PV, Hema A. Murthy

PS3.7 Instrument Identification in Optical Music Recognition
Yucong Jiang, Christopher Raphael

PS3.8 Cross-Version Singing Voice Detection in Classical Opera Recordings
Christian Dittmar, Bernhard Lehner, Thomas Prätzlich, Meinard Müller, Gerhard Widmer

PS3.9 Accurate Tempo Estimation based on Recurrent Neural Networks and Resonating Comb Filters
Sebastian Böck, Florian Krebs, Gerhard Widmer

PS3.10 Musicology of Early Music with Europeana Tools and Services
Erik Duval, Marnix van Berchum, Anja Jentzsch, Gonzalo Alberto Parra Chico, Andreas Drakos

PS3.11 Singing Voice Separation from Monaural Music Based on Kernel Back-Fitting Using Beta-Order Spectral Amplitude Estimation
Hye-Seung Cho, Jun-Yong Lee, Hyoung-Gook Kim

PS3.12 Schematizing the Treatment of Dissonance in 16th-Century Counterpoint
Andie Sigler, Jon Wild, Eliot Handelman

PS3.13 Analysis of the Evolution of Research Groups and Topics in the ISMIR Conference
Mohamed Sordo, Mitsunori Ogihara, Stefan Wuchty

PS3.14 A Comparison of Symbolic Similarity Measures for Finding Occurrences of Melodic Segments
Berit Janssen, Peter van Kranenburg, Anja Volk

PS3.15 PAD and SAS: Two Awareness-Weighted Rhythmic Similarity Distances
Daniel Gómez-Marín, Sergi Jordà, Perfecto Herrera

PS3.16 Four Timely Insights on Automatic Chord Estimation
Eric J. Humphrey, Juan P. Bello

PS3.17 Improving Melodic Similarity in Indian Art Music Using Culture-Specific Melodic Characteristics
Sankalp Gulati, Joan Serrà, Xavier Serra

PS3.18 Searching Lyrical Phrases in A-Capella Turkish Makam Recordings
Georgi Dzhambazov, Sertan Şentürk, Xavier Serra

PS3.19 Quantifying Lexical Novelty in Song Lyrics
Robert J. Ellis, Zhe Xing, Jiakun Fang, Ye Wang

PS3.20 An Efficient Temporally-Constrained Probabilistic Model for Multiple-Instrument Music Transcription
Emmanouil Benetos, Tillman Weyde

PS3.21 Electric Guitar Playing Technique Detection in Real-World Recordings Based on F0 Sequence Pattern Recognition
Yuan-Ping Chen, Li Su, Yi-Hsuan Yang

PS3.22 Extending a Model of Monophonic Hierarchical Music Analysis to Homophony
Phillip B. Kirlin, David L. Thomas

PS3.23 The MIR Perspective on the Evolution of Dynamics in Mainstream Music
Emmanuel Deruty, François Pachet

PS3.24 Theme and Variation Encodings with Roman Numerals (TAVERN): A New Data Set for Symbolic Music Analysis
Johanna Devaney, Claire Arthur, Nathaniel Condit-Schultz, Kirsten Nisula

PS3.25 Benford's Lawfor Music Analysis
Isabel Barbancho, Lorenzo J. Tardón, Ana M. Barbancho, Mateu Sbert

PS3.26 An Audio to Score Alignment Framework Using Spectral Factorization and Dynamic Time Warping
J.J. Carabias-Orti, F.J. Rodriguez-Serrano, P. Vera-Candeas, N. Ruiz-Reyes, F.J. Cañadas-Quesada

PS3.27 New Sonorities for Early Jazz Recordings Using Sound Source Separation and Automatic Mixing Tools
Daniel Matz, Estefanía Cano, Jakob Abeßer

PS3.28 Automatic Transcription of Ornamented Irish Traditional Flute Music Using Hidden Markov Models
Peter Jančovič, Münevver Köküer, Wrena Baptiste

PS3.29 Towards Music Imagery Information Retrieval: Introducing the OpenMIIR Dataset of EEG Recordings from Music Perception and Imagination
Sebastian Stober, Avital Sternin, Adrian M. Owen, Jessica A. Grahn

PS3.30 Emotion Based Segmentation of Musical Audio
Anna Aljanaki, Frans Wiering, Remco C. Veltkamp

11:30-12:30	**Keynote Speaker 2**
	The Promise of Music Information Retrieval: Are we there yet? *J. Stephen Downie*
12:30-13:30	**Industrial Panel** *Session Chair: Alexander Lerch*
13:30-14:30	Lunch
14:30-16:00	**Oral Session 6. User & Communnity** *Session Chair: Yi-Hsuan Yang*

OS6.1 MIREX Grand Challenge 2014 User Experience: Qualitative Analysis of User Feedback
Jin Ha Lee, Xiao Hu, Kahyun Choi, J. Stephen Downie

OS6.2 AcousticBrainz: A Community Platform for Gathering Music Information Obtained from Audio
Alastair Porter, Dmitry Bogdanov, Robert Kaye, Roman Tsukanov, Xavier Serra

OS6.3 How Music Alters Decision Making - Impact of Music Stimuli on Emotional Classification
Elad Liebman, Peter Stone, Corey N. White

OS6.4 Put the Concert Attendee in the Spotlight. A User-Centered Design and Development Approach for Classical Concert Applications
Mark S. Melenhorst, Cynthia C. S. Liem

16:00-17:30	**Poster Session 3** & Coffee
20:00	ISMIR 2015 Gala Dinner & Pandora Jam Session

Friday, October 30 Venue: Hotel NH Málaga

09:00-10:00 **Oral Session 7. Performance**
Session Chair: Zhiyao Duan

OS7.1 Analysis of Expressive Musical Terms in Violin Using Score-Informed and Expression-Based Audio Features
Pei-Ching Li, Li Su, Yi-Hsuan Yang, Alvin W. Y. Su

OS7.2 Spectral Learning for Expressive Interactive Ensemble Music Performance
Guangyu Xia, Yun Wang, Roger B. Dannenberg, Geoffrey Gordon

OS7.3 Score-Informed Analysis of Intonation and Pitch Modulation in Jazz Solos
Jakob Abeßer, Estefanía Cano, Klaus Frieler, Martin Pfleiderer, Wolf-Georg Zaddach

10:00-11:00 **MIREX Oral Session**
Session Chair: J. Stephen Downie

11:00-12:30 **MIREX Poster Session** & Coffee

12:30-13:30 **Business Meeting**
Session Chair: Simon Dixon

13:30-13:40 **Closing**

13:40-14:30 Lunch

14:30-16:00 **Demos & Late-Breaking News**

16:00-17:30 **Unconference**
Session Chairs: Daniel P. W. Ellis and Douglas Eck

WiMIR

WiMIR is a group of people dedicated to promoting the role of and increasing opportunities for women in the MIR field. We meet to socialize, share information, and discuss in an informal setting, with the goal of building a community around women, as well as men who support women, in our field. WiMIR has held an annual meeting at the ISMIR conference since 2012, having a high turnout of both female and male attendees. This year's meeting will include a short overview of the group and ongoing initiatives, and a mentorship session for ISMIR members.

Industrial Panel

The aim of the Industrial Panel is to help the researchers understand how industry is connected to research and academia, grants, internships, projects, etc.

Thus, the Industrial Panel will serve to connect more strongly the industry activities to the research community. The participants will have the opportunity to get from the panelists an idea of the main MIR/processing/management tasks they expect to be relevant for them in the next years.

Also, the audience will have the opportunity to freely ask some questions to the panel.

The panelists (in alphabetic order) will be representatives from the following enterprises: Pandora, Google, LyricFind, Steinberg and Smule.

MIREX

The Music Information Retrieval Evaluation eXchange (MIREX) is a collective effort to evaluate cutting-edge methods for various MIR tasks. Since 2005, MIREX has been an integral part of ISMIR. This year, ISMIR features the following MIREX-related events:

- Plenary summary of tasks and results of MIREX 2015
- Report of the "Grand Challenge 2015: User Experience" (GC15UX)
- Plenary discussion of MIREX 2016 and GC16UX
- Poster session of MIREX 2015 participants

Business Meeting

In this plenary session, the ISMIR board members present a summary of their work and activities to the community since last ISMIR and the forthcoming actions and decisions.

Late-Breaking and Demo (LBD)

Friday afternoon is dedicated to late-breaking contributions and MIR system demonstrations. Abstracts for these presentations are available online. The Late-Breaking and Demo (LBD) session aims to present preliminary results, ideas, applications or system prototypes that are not fully formed yet nor systematically evaluated, but of interest to the MIR community. As in previous editions, we will all enjoy brand new ideas and applications.

Unconference

Friday afternoon, we also have a special "Unconference" session. The Unconference will begin with a brief plenary to propose and vote on topics. Then, the participant will divide into groups to discuss on the topics of greatest interest. After an hour, all the participant will reunite at the plenary room, report back, and repeat for a second round.

The Unconference will be totally flexible featuring simple discussions among interested parties, short position talks followed by discussion, demos, etc.

A free-form Google doc has been created to foster ISMIR participants to contribute with ideas and comments:
https://docs.google.com/document/d/
1nP9TVcTC4rNEEGq9q38Ef8IBg43dX8qHvDb0e-eX_oc/edit?usp=sharing

Social Program

ISMIR 2015 not only offers interesting talks, papers, posters, and tutorials, but also aims at giving the participants an unforgettable stay. The social program provides participants with an opportunity to relax after meetings, to experience Málaga, and to network with other ISMIR participants. The social program includes:

Moday, October 26, 20:00	Welcome Reception *"Vinoteca Museo Los Patios de Beatas"* Calle Beatas 43, 29006 Málaga, Spain
Wednesday, October 28, 18:00	ISMIR 2015 Concert *"Sala de Conciertos María Cristina"* Calle Marqués de Valdecañas 2, 29008 Málaga, Spain
Thursday, October 29, 20:00	ISMIR 2015 Gala Dinner & Pandora Jam Session *"Hacienda del Alamo"* Camino de Casabermeja 130, 29014 Málaga, Spain (Bus transportation will be provided at 19:30 from *Hotel NH Málaga* to *"Hacienda del Alamo"* and back to *Hotel NH Málaga* after the ISMIR 2015 Gala Dinner & Pandora Jam Session)

ISMIR 2015 Concert

Wednesday, October 24, 18:00-21:00
Sala Unicaja de Conciertos María Cristina

Part I - Sinfonietta of San Francisco de Paula & PHENICX project

1.	Overture to The Creatures of Prometheus	L. v. Beethoven
2.	Suite Española, Op. 47 - "Sevilla"	Isaac Albéniz

Sinfonietta of San Francisco de Paula
Conductor Michael Thomas
Enriched by music visualization technology by PHENICX project

Part II - Flamenco Show

1.	Nostalgia de un malagueño	"Malagueña" dedicated to Picasso
2.	Tradiciones malagueñas	Sevillanas
3.	Guajira del Barquillero	Cantes de ida y vuelta
4.	La rosa del rosal	Malagueña por Verdiales
5.	Salve Rociera	Canción popular andaluza
6.	Málaga quita el sentío	Rumbas

"Coro Aire Andaluz"
Flamenco dancers Dolores Vargas and Mercedes Vargas

Part III - ISMIR 2015 Music

1.	Danza Española (Live electronics)	Takuro Shibayama Takayuki Hamano Hidefumi Ohmura
2.	Big Gizmo (fixed media electronic music 2 or 8 channels)	Sever Tipei
3.	Migrant (a cyclic piece for piano and computer)	Bill Manaris John Balafoukas
4.	Cante Reciclado (fixed media electronic music 2 channels)	Nadine Kroher
5.	Mai morente for Cello and tape	José Lopez-Montes
6.	Guided improvisation with SoundFisher (Live electronics)	Thom Blum
7.	Separation Logic, for trumpet and interactive system	Roger Dannenberg

Program Notes
Part I - Sinfonietta of San Francisco de Paula & PHENICX project

PHENICX: technologies for enriched concert experiences

An orchestra concert embraces a wealth of musical information, which may not be easily perceived or understood by general audiences. Current machine listening and visualization technologies can facilitate the appreciation of distinct musical facets, contributing to innovative and more enjoyable concert experiences. We provide an enriched experience for the Overture to The Creatures of Prometheus by L. v. Beethoven with four different visualizations:

1. Hear and see the SOUND
2. Follow the score of the PIECE in real-time
3. Understand the gestures of the CONDUCTOR
4. Visualize the activity of the ORCHESTRA sections

Some of these layers will be also shown during "Sevilla" movement of the "Spanish Suite" by Isaac Albéniz.

PHENICX (*phenicx.upf.edu*) is a new way of creating digital experiences for live performances of classical music. By integrating current music technologies, our goal is to enrich concert experiences and bring them to new audiences, providing novel engaging and interactive ways to explore live classical music. The PHENICX project (Feb 2013-Feb 2016) has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number 601166.

The Sinfonietta of San Francisco de Paula

This is a newly formed orchestra, created in 2011 by young musicians from Seville, Spain (*http://www.sinfoniettadesanfranciscodepaula.com/*). Its career has been very intense and it has given, among many others, concerts at the Teatro de la Maestranza and the Alcazar of Seville in 2012, performances in the Olympic Stadium in Seville and Burgos Cathedral in 2013. More recently the Sinfonietta has performed at the Basilica del Gran Poder de Sevilla with the organist Pedro Luengo, and has given an outstanding concert in the parish of St. José Obrero in Seville. In March 2015, it performed again at the Teatro de la Maestranza in Seville, at the opening concert of Singularity University Summit Spain.

For years now, The Sinfonietta is a guest orchestra in various and important events in France. In 2012 it had the opportunity to perform several concerts in Paris, organized by the cultural organization "Chevalier de Saint Georges".

In 2013 it participated as a guest orchestra at the international festival "Eurochestries", having the opportunity to play a common concert with other international orchestras. In 2015 The Sinfonietta was again invited to "Eurochestries" festival, and this time at its main and more demanding venue, the Pons headquarters, where it obtained an overwhelming success.

In October 2014, The Foundation Goñi and Rey incorporates the Sinfonietta of San Francisco de Paula in the institution. In such a way, it intends to promote one of the pillars of its fundamental purpose, culture, culture through music and education, providing an opportunity for young Sevillians to learn and share their talents with others, and to be on track toward achieving the maximum results. Getting beyond borders, both physical and abstract with already demonstrated proficiency and the name of Seville in each of its musical cases.

Michael Thomas (conductor)

Michael Thomas (born in Middlesbrough, UK, 1960) is a British violinist, composer and conductor. He is currently Chief Conductor of the Bética Chamber Orchestra, City of Almería Orchestra and the Baroque Youth Orchestra of Andalusia.

Program Notes
Part II - Flamenco Show

Coro Aire Andaluz

"Coro Aire Andaluz" is a flamenco chorus created in 1989. Its main objective is to promote the Spanish folklore and, specifically, the folklore from Málaga. This chorus is conducted by Isabel López Mayorga. Isabel has composed most of the pieces this chorus performs. "Coro Aire Andaluz" has performed mainly in Spain, Portugal and France, but also in other countries.

Dolores Vargas Jiménez (Flamenco dancers)

Dolores Vargas Jimenez 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.

Program Notes
Part III - ISMIR 2015 Music

The submissions to the ISMIR 2015 concert were very interesting and varied, demonstrating sophisticated uses of technology in the context of compelling musical expressions. Established composers include Sever Tipei (whose Big Gizmo uses an algorithmic generator that produces an entire family of compositions related to the one we will hear tonight), Thom Blum (using SoundFisher to query for sounds matching a seed, thereby improvising the result we hear), and Roger Dannenberg, a leader in this field whose acuity as a performer, composer, and developer (he pioneered the score-following techniques at work in this piece) blend to shape his Separation Logic. Nadine Kroher and José López-Montes both explore the flamenco tradition algorithmically with delightful and very different results. Shibayama, Hamano, and Ohmura also mine the rich tradition of Spanish music through a recombination of materials realized with Max/MSP. Finally, Bill Manaris and John Balafoukas present an eerily timely musical composition based on the movements of 56,976 immigrants to the United States.

Abstracts

Keynote Talks

Keynote Talk 1: 10:30-11:30 Tuesday, October 27

Integrating Music Information Sources for Music Production and Consumption

Mark Sandler, *Queen Mary University of London, UK*

For several years, research at Queen Mary's Centre for Digital Music has probed the intersection of signal analysis technologies with informatics technologies. More specifically, we have built audio signal-level feature extractors which output RDF-the language of the Semantic Web-which have enabled us to build prototypes that expose enhanced functionality and offer new experiences to music users of all kinds. This talk will summarise some of our early and recent work in Semantic Audio and Music Informatics, leading up to the current research themes of our FAST-IMPACt (Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption) project. FAST-IMPACt is a 5 year programme of research involving 3 UK universities together with commercial and non-commercial partners from around the world. Its overarching aim is to bring more engaging and immersive experiences based on musical knowledge of all kinds to users of all kinds. Where relevant and possible, the principles will be illustrated with demos.

Biography

Mark Sandler was born in 1955. He received the B.Sc. and Ph.D. degrees from the University of Essex, U.K., in 1978 and 1984, respectively. His PhD was an investigation into Digital Audio Power Amplification and he has been an active researcher in Digital Audio and Digital Music ever since. He is a Professor of Signal Processing at Queen Mary University of London, London, U.K., where he founded the Centre for Digital Music and he has published over 400 papers in journals and conferences and supervised over 30 PhD students. Mark Sandler is currently Director of the EPSRC/AHRC Centre for Doctoral Training in Media and Arts Technology, and Principal Investigator of the 5 year research project, Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption. Mark is a Fellow of the Institute of Engineering and Technol-

ogy (FIET), a Fellow of the Audio Engineering Society (FAES), a Fellow of the British Computer Society (FBCS), and a Fellow of the Institution of Electronic and Electrical Engineers (FIEEE).

Keynote Talk 2: **11:30-12:30 Thursday, October 29**

The Promise of Music Information Retrieval: Are we there yet?

J. Stephen Downie, *University of Illinois at Urbana-Champaign, USA*

In 1988 I needed to find some easy-to-play music for an upcoming flute performance exam. As a famously mediocre flute player, I was more than a little desperate to find something that would require the least amount of rehearsal time (and talent) to perform. I had previously played a baroque piece in the key of C that just might fit the bill. After too many hours of searching that included innumerable consultations with infinitely patient music librarians and fellow music students, we finally determined that the work in question was Bach's Flute Sonata in C, BWV 1033. This real-world use case inspired me to ask: "Why is it so difficult to identify, and then locate, a rather famous piece of music that I had actually played before?" In this talk, I reflect on just how far music information retrieval research (MIR) has come in the intervening 27 years in making access to music resources quick and simple. As I expound on my personal journey as an MIR researcher, I will note some of the twists in the road that I have surprised me. I will also offer up some commentary on lesser-travelled paths that we as community should explore.

Biography J. Stephen Downie is the Associate Dean for Research and a Professor at the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. Downie is the Illinois Co-Director of the HathiTrust Research Center (HTRC). He is also Director of the International Music Information Retrieval Systems Evaluation Laboratory (IMIRSEL) and founder and ongoing director of the Music Information Retrieval Evaluation eXchange (MIREX). Stephen Downie was the Principal Investigator on the Networked Environment for Music Analysis (NEMA) project, funded by the Andrew W. Mellon Foundation. Furthermore, he is Co-PI on the Structural Analysis of Large Amounts of Music Information (SALAMI) project, jointly funded by the National Science Foundation (NSF), the Canadian Social Science and Humanities Research Council (SSHRC), and the UK's Joint Information Systems Committee (JISC). Stephen Downie has been very active in the establishment of the Music Information Retrieval (MIR) community through his ongoing work with the International Society for Music Information Retrieval (IS-

MIR) conferences. He was ISMIR's founding President and now serves on the ISMIR board. Professor Downie holds a BA (Music Theory and Composition) along with a Master's and a PhD in Library and Information Science, all earned at the University of Western Ontario, London, Canada.

Tutorials

Tutorial 1: 10:00-13:30 Monday, October 26

Why Singing is Interesting

Simon Dixon, *Queen Mary University of London, UK*
Masataka Goto, *AIST, Japan*
Matthias Mauch, *Queen Mary University of London, UK*

This tutorial aims to introduce to the ISMIR community the exciting world of singing styles, the mechanisms of the singing voice, and provide a guide to representations, engineering tools, and methods for analyzing and leveraging it. The singing voice is arguably the most expressive of all musical instruments, and all popular music cultures around the world use singing. Across disciplines, a lot is known about singing culture and the intricate physiological and psychological mechanisms of singing, but this knowledge is not exploited enough in much of the music information retrieval literature. The three parts of the tutorial (one hour each) are designed to remedy this: an introduction to singing styles, techniques and forms around the world (including a short introduction to the psychology of singing), a practical guide to the analysis of singing using music informatics tools, and an overview over various systems for singing information processing. Our aim is for music information retrieval specialists to walk away with a newly sparked passion for singing, and ideas of how to use our knowledge of singing, and singing information processing, to create new, exciting research.

Tutorial 2: 10:00-13:30 Monday, October 26

Addressing the Music Information Needs of Musicologists

Richard J. Lewis, *Goldsmiths College, University of London, UK*
Ben Fields, *Goldsmiths College, University of London, UK*
Tim Crawford, *Goldsmiths College, University of London, UK*

The music information needs of musicologists are not being met by the current generation of MIR tools and techniques. While evaluation has always been central to the practice of the music information retrieval community, the tasks tackled most often address the music information needs of recreational users, such as playlist recommendation systems; or are specified at a level which is not very relevant to the needs of music researchers, such as beat or key finding; or have focused on-and possibly even become over-fitted to-a narrow range of musical repertoire which doesn't cover musicological interests. In this tutorial we will present those music information needs through topics including at least the following: the metadata requirements of historical musicology; working with symbolic corpora; studying musical networks; passage-level audio search; and musical understandings of audio features. As well as these scheduled presentations and discussions, we will ask the attendees to submit suggestions of musicologically motivated research questions suitable for MIR during the course of the tutorial. These will then be reviewed and discussed during the conclusion of the tutorial. Finally, we have invited Meinard Müller to conclude the tutorial by outlining his view on the current state of MIR for musicology. We are aiming to enable attendees, as experts in their own areas of MIR, to find new applications of their tools and techniques that can also serve the needs of musicologists. Given the selection of MIR topics we intend to cover, this tutorial will be of particular interest to those working in: musical metadata; symbolic MIR; audio search; and graph analytics. We believe contemporary musicology to be a rich source of new and exciting challenges for MIR and we are confident the community can rise to those challenges. In the long term, we hope this tutorial will give rise to a selection of new MIREX tasks that focus on musicological challenges.

Tutorial 3: **10:00-13:30 Monday, October 26**

Markov Logic Networks for Music Analysis

Helene Papadopoulos, *CNRS, Paris, France*

The automatic extraction of relevant content information from music audio signals is an essential aspect of Music Information Retrieval (MIR). Music audio signals are very rich and complex, both because of the intrinsic physical nature of audio (incomplete and noisy observations, many modes of sound production, etc.), and because they convey multi-faceted and strongly interrelated semantic information (harmony, melody, metric, structure, etc.). Dealing with real audio recordings thus requires the ability to handle both uncertainty and complex relational structure at multiple levels of representation. Until recent years, these two aspects have been generally treated separately, probability being the standard way to represent uncertainty in knowledge, while logical representation being used to represent complex relational information. Markov Logic Networks (MLNs), in which statistical and relational knowledge are unified within a single representation formalism, have recently received considerable attention in many domains such as natural language processing, link-based Web search, or bioinformatics. The goal of this tutorial is to provide a comprehensive overview of Markov logic networks and show how they can be used as a highly flexible and expressive yet concise formalism for the analysis of music audio signals. We will show how MLNs encompass the probabilistic and logic-based models that are classically used in MIR. Algorithms for MLN modeling, training and inference will be presented, as well as open-source software packages for MLNs that are suitable to MIR applications. We will discuss concrete case-study examples in various fields of application.

Tutorial 4: 14:30-18:00 Monday, October 26

COmputation and FLAmenco: Why Flamenco is Interesting for MIR Research

Emilia Gómez, *Universitat Pompeu Fabra, Barcelona, Spain*
Nadine Kroher, *Universitat Pompeu Fabra, Barcelona, Spain*
José Miguel Díaz-Bañez, *Universidad de Sevilla, Spain*
Sergio Oramas, *Universitat Pompeu Fabra, Barcelona, Spain*
Joaquín Mora, *Universidad de Sevilla, Spain*
Francisco Gómez-Martín, *Universidad Politécnica de Madrid, Spain*

This tutorial provides an introduction to flamenco music with the support of MIR techniques. At the same time, the tutorial analyzes the challenges and opportunities that this music repertoire offers MIR researchers, presents some research contributions and provides a forum to discuss about how to address those challenges in future research. As ISMIR 2015 is in Málaga, this tutorial will give ISMIR participants a unique chance to discover flamenco music in its original location. The tutorial will be structured in two main parts. First, we will provide a general introduction to flamenco music: origins and evolution, musical characteristics, instrumentation, singing and guitar. We will illustrate this introduction with multimedia material and live performance. Then we will analyze how MIR technologies perform for flamenco music. By discussing several MIR tasks and how they should be addressed in this context, we will discover more about flamenco and how methods tailored to this repertoire can be exploited in other contexts. We will focus on automatic transcription, singer identification, music similarity, genre classification, rhythmic and melodic pattern detection and context-based music description methods. Participants will have the chance to interact with MIR annotated datasets and tools developed for flamenco music in the context of the COFLA project.

Tutorial 5: **14:30-18:00 Monday, October 26**

Using Correlation Analysis and Big Data to Identify and Predict Musical Behaviors

Jeff C. Smith, *Smule*

New and significant repositories of musical data afford unique opportunities to apply data analysis techniques to ascertain insights of musical engagement. These repositories include performance, listening, curation, and behavioral data. Often the data in these repositories also includes demographic and/or location information, allowing studies of musical behavior, for example, to be correlated with culture or geography. Historically, the analysis of musical behaviors was limited. Often, subjects (e.g. performers or listeners) were recruited for such studies. This technique suffered from issues around methodology (e.g. the sample set of subjects would often exhibit bias) or an insufficient number of subjects and/or data to make reliable statements of significance. That is to say the conclusions from these studies were largely anecdotal. In contrast to these historical studies, the availability of new repositories of musical data allow for studies in musical engagement to develop conclusions that pass standards of significance, thereby yielding actual insights into musical behaviors. This tutorial will demonstrate several techniques and examples where correlation and statistical analysis is applied to large repositories of musical data to document various facets of musical engagement. Web site: https://ccrma.stanford.edu/damp/ Stanford University has created a new corpus of amateur music performance data, the Stanford Digital Archive of Mobile Performances, or DAMP, to facilitate the study of musical engagement through application of correlation and statistical analysis.

Tutorial 6: 14:30-18:00 Monday, October 26

Automatic Music Transcription

Zhiyao Duan, *University of Rochester, USA*
Emmanouil Benetos, *Queen Mary University of London, UK*

Automatic Music Transcription (AMT) is a fundamental problem in music information retrieval. Roughly speaking, transcription refers to extracting a symbolic representation —a list of notes (pitches and rhythms)—from an audio signal. Music transcription is a fascinating but challenging task, even for humans: in undergraduate music education it is usually called dictation, and achieving a high level of proficiency requires years of practice and training. Empowering machines with this ability is an even more challenging problem, especially for automatically transcribing polyphonic music. To that end, the AMT problem has drawn great interest of researchers from several areas including signal processing, machine learning, acoustics, music theory, and music cognition. In terms of applications, a successful AMT system would be helpful for solving many MIR research problems, including music source separation, structure analysis, content-based music retrieval, and musicological study of non-notated music, just to name a few. This tutorial will give an overview of the AMT problem, including current approaches, datasets and evaluation methodologies. It will also explore connections with other related problems (i.e. audio-score alignment, source separation) as well as applications to related fields, such as content-based music retrieval and computational musicology. The tutorial is designed for students and researchers who have general knowledge of music information retrieval and/or computational musicology and are interested in getting into the field of AMT. A substantial amount of time will be spent in discussing challenges and research directions; we hope that this discussion will help move this field forward, and influence related fields in MIR and computational musicology to exploit AMT technologies. The tutorial will also include hands-on sessions on using AMT code and plugins - participants will be encouraged to bring their laptops and gain access to transcription datasets, as well as work on AMT examples.

Papers

Poster Session 1 11:30-13:30, 16:00-17:30 Tuesday, October 27

PS1.1 Image Quality Estimation for Multi-Score OMR

Dan Ringwalt, Roger B. Dannenberg

Optical music recognition (OMR) is the recognition of images of musical scores. Recent research has suggested aligning the results of OMR from multiple scores of the same work (multi-score OMR, MS-OMR) to improve accuracy. As a simpler alternative, we have developed features which predict the quality of a given score, allowing us to select the highest-quality score to use for OMR. Furthermore, quality may be used to weight each score in an alignment, which should improve existing systems' robustness. Using commercial OMR software on a test set of MIDI recordings and multiple corresponding scores, our predicted OMR accuracy is weakly but significantly correlated with the true accuracy. Improved features should be able to produce highly consistent results.

PS1.2 Comparative Music Similarity Modelling Using Transfer Learning Across User Groups

Daniel Wolff, Andrew MacFarlane, Tillman Weyde

We introduce a new application of transfer learning for training and comparing music similarity models based on relative user data: The proposed Relative Information-Theoretic Metric Learning (RITML) algorithm adapts a Mahalanobis distance using an iterative application of the ITML algorithm, thereby extending it to relative similarity data. RITML supports transfer learning by training models with respect to a given template model that can provide prior information for regularisation. With this feature we use information from larger datasets to build better models for more specific datasets, such as user groups from different cultures or of different age. We then evaluate what model parameters, in this case acoustic features, are relevant for the specific models when compared to the general user data. We to this end introduce the new CASimIR dataset, the first openly available relative similarity dataset with user attributes. With two age-related subsets, we show that transfer learning with RITML leads to better age-specific models. RITML here improves learning on small datasets. Using the larger MagnaTagATune dataset, we show that RITML performs as well as state-of-the-art algorithms in terms of general similarity estimation.

PS1.3 Modeling Genre with the Music Genome Project: Comparing Human-Labeled Attributes and Audio Features

Matthew Prockup, Andreas F. Ehmann, Fabien Gouyon, Erik M. Schmidt, Oscar Celma, Youngmoo E. Kim

Genre provides one of the most convenient categorizations of music, but it is often regarded as a poorly defined or largely subjective musical construct. In this work, we provide evidence that musical genres can to a large extent be objectively modeled via a combination of musical attributes. We employ a data-driven approach utilizing a subset of 48 hand-labeled musical attributes comprising instrumentation, timbre, and rhythm across more than one million examples from Pandora® Internet Radio's *Music Genome Project®*. A set of audio features motivated by timbre and rhythm are then implemented to model genre both directly and through audio-driven models derived from the hand-labeled musical attributes. In most cases, machine learning models built directly from hand-labeled attributes outperform models based on audio features. Among the audio-based models, those that combine audio features and learned musical attributes perform better than those derived from audio features alone.

PS1.4 Cover Song Identification with Timbral Shape Sequences

Christopher J. Tralie, Paul Bendich

We introduce a novel low level feature for identifying cover songs which quantifies the relative changes in the smoothed frequency spectrum of a song. Our key insight is that a sliding window representation of a chunk of audio can be viewed as a time-ordered point cloud in high dimensions. For corresponding chunks of audio between different versions of the same song, these point clouds are approximately rotated, translated, and scaled copies of each other. If we treat MFCC embeddings as point clouds and cast the problem as a relative shape sequence, we are able to correctly identify 42/80 cover songs in the "Covers 80" dataset. By contrast, all other work to date on cover songs exclusively relies on matching note sequences from Chroma derived features.

PS1.5 On the Impact of Key Detection Performance for Identifying Classical Music Styles

Christof Weiß, Maximilian Schaab

We study the automatic identification of Western classical music styles by directly using chroma histograms as classification features. Thereby, we evaluate the benefits of knowing a piece's global key for estimating key-related pitch classes. First, we present four automatic key detection systems. We compare their performance on suitable datasets of classical music and optimize the algorithms' free parameters. Using a second dataset, we evaluate automatic classifi-

cation into the four style periods Baroque, Classical, Romantic, and Modern. To that end, we calculate global chroma statistics of each audio track. We then split up the tracks according to major and minor keys and circularly shift the chroma histograms with respect to the tonic note. Based on these features, we train two individual classifier models for major and minor keys. We test the efficiency of four chroma extraction algorithms for classification. Furthermore, we evaluate the impact of key detection performance on the classification results. Additionally, we compare the key-related chroma features to other chroma-based features. We obtain improved performance when using an efficient key detection method for shifting the chroma histograms.

PS1.6 Chord Detection Using Deep Learning

Xinquan Zhou, Alexander Lerch

In this paper, we utilize deep learning to learn high-level features for audio chord detection. The learned features, obtained by a deep network in bottleneck architecture, give promising results and outperform state-of-the-art systems. We present and evaluate the results for various methods and configurations, including input pre-processing, a bottleneck architecture, and SVMs vs. HMMs for chord classification.

PS1.7 Temporal Music Context Identification with User Listening Data

Cameron Summers, Phillip Popp

The times when music is played can indicate context for listeners. From the peaceful song for waking up each morning to the traditional song for celebrating a holiday to an up-beat song for enjoying the summer, the relationship between the music and the temporal context is clearly important. For music search and recommendation systems, an understanding of these relationships provides a richer environment to discover and listen. But with the large number of tracks available in music catalogues today, manually labeling track-temporal context associations is difficult, time consuming, and costly. This paper examines track-day contexts with the purpose of identifying relationships with specific music tracks. Improvements are made to an existing method for classifying Christmas tracks and a generalization to the approach is shown that allows automated discovery of music for any day of the year. Analyzing the top 50 tracks obtained from this method for three well-known holidays, Halloween, Saint Patrick's Day, and July 4th, precision@ 50 was 95%, 99%, and 73%, respectively.

PS1.8 Improving Music Recommendations with a Weighted Factorization of the Tagging Activity

Andreu Vall, Marcin Skowron,Peter Knees, Markus Schedl

Collaborative filtering systems for music recommendations are often based on implicit feedback derived from listening activity. Hybrid approaches further incorporate additional sources of information in order to improve the quality of the recommendations. In the context of a music streaming service, we present a hybrid model based on matrix factorization techniques that fuses the implicit feedback derived from the users' listening activity with the tags that users have given to musical items. In contrast to existing work, we introduce a novel approach to exploit tags by performing a weighted factorization of the tagging activity. We evaluate the model for the task of artist recommendation, using the expected percentile rank as metric, extended with confidence intervals to enable the comparison between models. Thus, our contribution is twofold: (1) we introduce a novel model that uses tags to improve music recommendations and (2) we extend the evaluation methodology to compare the performance of different recommender systems.

PS1.9 An Efficient State-Space Model for Joint Tempo and Meter Tracking

Florian Krebs, Sebastian Böck, Gerhard Widmer

Dynamic Bayesian networks (e.g., Hidden Markov Models) are popular frameworks for meter tracking in music because they are able to incorporate prior knowledge about the dynamics of rhythmic parameters (tempo, meter, rhythmic patterns, etc.). One popular example is the bar pointer model, which enables joint inference of these rhythmic parameters from a piece of music. While this allows the mutual dependencies between these parameters to be exploited, it also increases the computational complexity of the models. In this paper, we propose a new state-space discretisation and tempo transition model for this class of models that can act as a drop-in replacement and not only increases the beat and downbeat tracking accuracy, but also reduces time and memory complexity drastically. We incorporate the new model into two state-of-the-art beat and meter tracking systems, and demonstrate its superiority to the original models on six datasets.

PS1.10 Automatic Handwritten Mensural Notation Interpreter: From Manuscript to MIDI Performance

Yu-Hui Huang, Xuanli Chen, Serafina Beck, David Burn, Luc Van Gool

This paper presents a novel automatic recognition framework for hand-written mensural music. It takes a scanned manuscript as input and yields as output modern music scores. Compared to the previous mensural Optical Music Recognition (OMR) systems, ours shows not only promising performance in music recogni-

tion, but also works as a complete pipeline which integrates both recognition and transcription.
There are three main parts in this pipeline: i) region-ofinterest detection, ii) music symbol detection and classification, and iii) transcription to modern music. In addition to the output in modern notation, our system can generate a MIDI file as well. It provides an easy platform for the musicologists to analyze old manuscripts. Moreover, it renders these valuable cultural heritage resources available to non-specialists as well, as they can now access such ancient music in a better understandable form.

PS1.11 Infinite Superimposed Discrete All-Pole Modeling for Multipitch Analysis of Wavelet Spectrograms

Kazuyoshi Yoshii, Katsutoshi Itoyama, Masataka Goto

This paper presents a statistical multipich analyzer based on a source-filter model that decomposes a target music audio signal in terms of three major kinds of sound quantities: pitch (fundamental frequency: F0), timbre (spectral envelope), and intensity (amplitude). If the spectral envelope of an isolated sound is represented by an all-pole filter, linear predictive coding (LPC) can be used for filter estimation in the linear-frequency domain. The main problem of LPC is that although only the amplitudes of harmonic partials are reliable samples drawn from the spectral envelope, the whole spectrum is used for filter estimation. To solve this problem, we propose an *infinite superimposed discrete all-pole* (iSDAP) model that, given a music signal, can estimate an appropriate number of superimposed harmonic structures whose harmonic partials are drawn from a limited number of spectral envelopes. Our nonparametric Bayesian source-filter model is formulated in the logfrequency domain that better suits the frequency characteristics of human audition. Experimental results showed that the proposed model outperformed the counterpart model formulated in the linear frequency domain.

PS1.12 Melodic Similarity in Traditional French-Canadian Instrumental Dance Tunes

Laura Risk, Lillio Mok, Andrew Hankinson, Julie Cumming

Commercial recordings of French-Canadian instrumental dance tunes represent a varied and complex corpus of study. This was a primarily aural tradition, transmitted from performer to performer with few notated sources until the late 20th century. Practitioners routinely combined tune segments to create new tunes and personalized settings of existing tunes. This has resulted in a corpus that exhibits an extreme amount of variation, even among tunes with the same name. In addition, the same tune or tune segment may appear under several different names. Previous attempts at building systems for automated retrieval and ranking of in-

strumental dance tunes perform well for near-exact matching of tunes, but do not work as well in retrieving and ranking, in order of most to least similar, variants of a tune; especially those with variations as extreme as this particular corpus. In this paper we will describe a new approach capable of ranked retrieval of variant tunes, and demonstrate its effectiveness on a transcribed corpus of incipits.

PS1.13 A Semantic-Based Approach for Artist Similarity

Sergio Oramas, Mohamed Sordo, Luis Espinosa-Anke, Xavier Serra

This paper describes and evaluates a method for computing artist similarity from a set of artist biographies. The proposed method aims at leveraging semantic information present in these biographies, and can be divided in three main steps, namely: (1) entity linking, i.e. detecting mentions to named entities in the text and linking them to an external knowledge base; (2) deriving a knowledge representation from these mentions in the form of a semantic graph or a mapping to a vector-space model; and (3) computing semantic similarity between documents. We test this approach on a corpus of 188 artist biographies and a slightly larger dataset of 2,336 artists, both gathered from Last.fm. The former is mapped to the MIREX Audio and Music Similarity evaluation dataset, so that its similarity judgments can be used as ground truth. For the latter dataset we use the similarity between artists as provided by the Last.fm API. Our evaluation results show that an approach that computes similarity over a graph of entities and semantic categories clearly outperforms a baseline that exploits word co-occurrences and latent factors.

PS1.14 Predicting Pairwise Pitch Contour Relations Based on Linguistic Tone Information in Beijing Opera Singing

Shuo Zhang, Rafael Caro Repetto, Xavier Serra

The similarity between linguistic tones and melodic pitch contours in Beijing Opera can be captured either by the contour shape of single syllable units, or by the pairwise pitch height relations in adjacent syllable units. In this paper, we investigate the latter problem with a novel machine learning approach, using techniques from time series data mining. Approximately 1300 pairwise contour segments are extracted from a selection of 20 arias. We then formulate the problem as a supervised machinelearning task of predicting types of pairwise melodic relations based on linguistic tone information. The results give a comparative view of fixed and mixed-effects models that achieved around 70% of maximum accuracy. We discuss the superiority of the current method to that of the unsupervised learning in single-syllable-unit contour analysis of similarity in Beijing Opera.

PS1.15 Song2Quartet: A System for Generating String Quartet Cover Songs from Polyphonic Audio of Popular Music

Graham Percival, Satoru Fukayama, Masataka Goto

We present Song2Quartet, a system for generating string quartet versions of popular songs by combining probabilistic models estimated from a corpus of symbolic classical music with the target audio file of any song. Song2Quartet allows users to add novelty to listening experience of their favorite songs and gain familiarity with string quartets. Previous work in automatic arrangement of music only used symbolic scores to achieve a particular musical style; our challenge is to also consider audio features of the target popular song. In addition to typical audio music content analysis such as beat and chord estimation, we also use time-frequency spectral analysis in order to better reflect partial phrases of the song in its cover version. Song2Quartet produces a probabilistic network of possible musical notes at every sixteenth note for each accompanying instrument of the quartet by combining beats, chords, and spectrogram from the target song with Markov chains estimated from our corpora of quartet music. As a result, the musical score of the cover version can be generated by finding the optimal paths through these networks. We show that the generated results follow the conventions of classical string quartet music while retaining some partial phrases and chord voicings from the target audio.

PS1.16 Exploring Data Augmentation for Improved Singing Voice Detection with Neural Networks

Jan Schlüter, Thomas Grill

In computer vision, state-of-the-art object recognition systems rely on label-preserving image transformations such as scaling and rotation to augment the training datasets. The additional training examples help the system to learn invariances that are difficult to build into the model, and improve generalization to unseen data. To the best of our knowledge, this approach has not been systematically explored for music signals. Using the problem of singing voice detection with neural networks as an example, we apply a range of label-preserving audio transformations to assess their utility for music data augmentation. In line with recent research in speech recognition, we find pitch shifting to be the most helpful augmentation method. Combined with time stretching and random frequency filtering, we achieve a reduction in classification error between 10 and 30%, reaching the state of the art on two public datasets. We expect that audio data augmentation would yield significant gains for several other sequence labelling and event detection tasks in music information retrieval.

PS1.17 Audio Chord Recognition with a Hybrid Recurrent Neural Network
Siddharth Sigtia, Nicolas Boulanger-Lewandowski, Simon Dixon

In this paper, we present a novel architecture for audio chord estimation using a hybrid recurrent neural network. The architecture replaces hidden Markov models (HMMs) with recurrent neural network (RNN) based language models for modelling temporal dependencies between chords. We demonstrate the ability of feed forward deep neural networks (DNNs) to learn discriminative features directly from a time-frequency representation of the acoustic signal, eliminating the need for a complex feature extraction stage. For the hybrid RNN architecture, inference over the output variables of interest is performed using beam search. In addition to the hybrid model, we propose a modification to beam search using a hash table which yields improved results while reducing memory requirements by an order of magnitude, thus making the proposed model suitable for real-time applications. We evaluate our model's performance on a dataset with publicly available annotations and demonstrate that the performance is comparable to existing state of the art approaches for chord recognition.

PS1.18 Design and Evaluation of a Probabilistic Music Projection Interface
Beatrix Vad, Daniel Boland, John Williamson, Roderick Murray-Smith, Peter Berg Steffensen

We describe the design and evaluation of a probabilistic interface for music exploration and casual playlist generation. Predicted subjective features, such as mood and genre, inferred from low-level audio features create a 34- dimensional feature space. We use a nonlinear dimensionality reduction algorithm to create 2D music maps of tracks, and augment these with visualisations of probabilistic mappings of selected features and their uncertainty. We evaluated the system in a longitudinal trial in users' homes over several weeks. Users said they had fun with the interface and liked the casual nature of the playlist generation. Users preferred to generate playlists from a local neighbourhood of the map, rather than from a trajectory, using neighbourhood selection more than three times more often than path selection. Probabilistic highlighting of subjective features led to more focused exploration in mouse activity logs, and 6 of 8 users said they preferred the probabilistic highlighting mode.

PS1.19 Conceptual Blending in Music Cadences: A Formal Model and Subjective Evaluation
Asterios Zacharakis, Maximos Kaliakatsos-Papakostas, Emilios Cambouropoulos

Conceptual blending is a cognitive theory whereby elements from diverse, but structurally-related, mental spaces are 'blended' giving rise to new conceptual

spaces. This study focuses on structural blending utilising an algorithmic formalisation for conceptual blending applied to harmonic concepts. More specifically, it investigates the ability of the system to produce meaningful blends between harmonic cadences, which arguably constitute the most fundamental harmonic concept. The system creates a variety of blends combining elements of the penultimate chords of two input cadences and it further estimates the expected relationships between the produced blends. Then, a preliminary subjective evaluation of the proposed blending system is presented. A pairwise dissimilarity listening test was conducted using original and blended cadences as stimuli. Subsequent multidimensional scaling analysis produced spatial configurations for both behavioural data and dissimilarity estimations by the algorithm. Comparison of the two configurations showed that the system is capable of making fair predictions of the perceived dissimilarities between the blended cadences. This implies that this conceptual blending approach is able to create perceptually meaningful blends based on self-evaluation of its outcome.

PS1.20 Harmonic-Percussive Source Separation Using Harmonicity and Sparsity Constraints

Jeongsoo Park, Kyogu Lee

In this paper, we propose a novel approach to harmonicpercussive sound separation (HPSS) using Non-negative Matrix Factorization (NMF) with sparsity and harmonicity constraints. Conventional HPSS methods have focused on temporal continuity of harmonic components and spectral continuity of percussive components. However, it may not be appropriate to use them to separate time-varying harmonic signals such as vocals, vibratos, and glissandos, as they lack in temporal continuity. Based on the observation that the spectral distributions of harmonic and percussive signals differ – i.e., harmonic components have harmonic and sparse structure while percussive components are broadband – we propose an algorithm that successfully separates the rapidly time-varying harmonic signals from the percussive ones by imposing different constraints on the two groups of spectral bases. Experiments with real recordings as well as synthesized sounds show that the proposed method outperforms the conventional methods.

PS1.21 A Hierarchical Bayesian Framework for Score-Informed Source Separation of Pianomusic Signals

Wai Man Szeto, Kin Hong Wong

Here we propose a score-informed monaural source separation system to extract every tone from a mixture of piano tone signals. Two sinusoidal models in our earlier work are employed in the above-mentioned system to represent piano tones: the General Model and the Piano Model. The General Model, a variant of sinusoidal modeling, can represent a single tone with high modeling

quality, yet it fails to separate mixtures of tones due to the overlapping partials. The PianoModel, on the other hand, is an instrument-specific model tailored for piano. Its modeling quality is lower but it can learn from training data (consisting entirely of isolated tones), resolve the overlapping partials and thus separate the mixtures. We formulate a new hierarchical Bayesian framework to run both Models in the source separation process so that the mixtures with overlapping partials can be separated with high quality. The results show that our proposed system gives robust and accurate separation of piano tone signal mixtures (including octaves) while achieving significantly better quality than those reported in related work done previously.

PS1.22 Automatic Tune Family Identification by Musical Sequence Alignment

Patrick E. Savage, Quentin D. Atkinson

Musics, like languages and genes, evolve through a process of transmission, variation, and selection. Evolution of musical tune families has been studied qualitatively for over a century, but quantitative analysis has been hampered by an inability to objectively distinguish between musical similarities that are due to chance and those that are due to descent from a common ancestor. Here we propose an automated method to identify tune families by adapting genetic sequence alignment algorithms designed for automatic identification and alignment of protein families. We tested the effectiveness of our method against a high-quality ground-truth dataset of 26 folk tunes from four diverse tune families (two English, two Japanese) that had previously been identified and aligned manually by expert musicologists. We tested different combinations of parameters related to sequence alignment and to modeling of pitch, rhythm, and text to find the combination that best matched the ground-truth classifications. The best-performing automated model correctly grouped 100% (26/26) of the tunes in terms of overall similarity to other tunes, identifying 85% (22/26) of these tunes as forming distinct tune families. The success of our approach on a diverse, cross-cultural ground-truth dataset suggests promise for future automated reconstruction of musical evolution on a wide scale.

PS1.23 Evaluation of Album Effect for Feature Selection in Music Genre Recognition

Igor Vatolkin, Günter Rudolph, Claus Weihs

With an increasing number of available music characteristics, feature selection becomes more important for various categorisation tasks, helping to identify relevant features and remove irrelevant and redundant ones. Another advantage is the decrease of runtime and storage demands. However, sometimes feature selection may lead to “overoptimisation” when data in the optimisation set is too

different from data in the independent validation set. In this paper, we extend our previous work on feature selection for music genre recognition and focus on so-called "album effect" meaning that optimised classification models may overemphasize relevant characteristics of particular artists and albums rather than learning relevant properties of genres. For that case we examine the performance of classification models on two validation sets after the optimisation with feature selection: the first set with tracks not used for training and feature selection but randomly selected from the same albums, and the second set with tracks selected from other albums. As it can be expected, the classification performance on the second set decreases. Nevertheless, in almost all cases the feature selection remains beneficial compared to complete feature sets and a baseline using MFCCs, if applied for an ensemble of classifiers, proving robust generalisation performance.

PS1.24 Music Pattern Discovery with Variable Markov Oracle: A Unified Approach to Symbolic and Audio Representations

Cheng-i Wang, Jennifer Hsu, Shlomo Dubnov

This paper presents a framework for automatically discovering patterns in a polyphonic music piece. The proposed framework is capable of handling both symbolic and audio representations. Chroma features are post-processed with heuristics stemming from musical knowledge and fed into the pattern discovery framework. The pattern-finding algorithm is based on *Variable Markov Oracle*. The *Variable Markov Oracle* data structure is capable of locating repeated suffixes within a time series, thus making it an appropriate tool for the pattern discovery task. Evaluation of the proposed framework is performed on the JKU Patterns Development Dataset with state of the art performance.

PS1.25 Automatic Solfège Assessment

Rodrigo Schramm, Helena de Souza Nunes, Cláudio Rosito Jung

This paper presents a note-by-note approach for automatic solfège assessment. The proposed system uses melodic transcription techniques to extract the sung notes from the audio signal, and the sequence of melodic segments is subsequently processed by a two stage algorithm. On the first stage, an aggregation process is introduced to perform the temporal alignment between the transcribed melody and the music score (ground truth). This stage implicitly aggregates and links the best combination of the extracted melodic segments with the expected note in the ground truth. On the second stage, a statistical method is used to evaluate the accuracy of each detected sung note. The technique is implemented using a Bayesian classifier, which is trained using an audio dataset containing individual scores provided by a committee of expert listeners. These individual scores were measured at each musical note, regarding the pitch, onset, and offset

accuracy. Experimental results indicate that the classification scheme is suitable to be used as an assessment tool, providing useful feedback to the student.

PS1.26 Evaluating Conflict Management Mechanisms for Online Social Jukeboxes

Felipe Vieira, Nazareno Andrade

Social music listening is a prevalent and often fruitful experience. Social jukeboxes are systems that enable social music listening with listeners collaboratively choosing the music to be played. Naturally, because music tastes are diverse, using social jukeboxes often involves conflicting interests. Because of that, virtually all social jukeboxes incorporate conflict management mechanisms. In contrast with their widespread use, however, little attention has been given to evaluating how different conflict management mechanisms function to preserve the positive experience of music listeners. This paper presents an experiment with three conflict management mechanisms and three groups of listeners. The mechanisms were chosen to represent those most commonly used in the state of the practice. Our study employs a mixed-methods approach to quantitatively analyze listeners' satisfaction and to examine their impressions and views on conflict, conflict management mechanisms, and social jukeboxing.

PS1.27 Particle Filters for Efficient Meter Tracking with Dynamic Bayesian Networks

Ajay Srinivasamurthy, Andre Holzapfel, Ali Taylan Cemgil, Xavier Serra

Recent approaches in meter tracking have successfully applied Bayesian models. While the proposed models can be adapted to different musical styles, the applicability of these flexible methods so far is limited because the application of exact inference is computationally demanding. More efficient approximate inference algorithms using particle filters (PF) can be developed to overcome this limitation. In this paper, we assume that the type of meter of a piece is known, and use this knowledge to simplify an existing Bayesian model with the goal of incorporating a more diverse observation model. We then propose Particle Filter based inference schemes for both the original model and the simplification. We compare the results obtained from exact and approximate inference in terms of meter tracking accuracy as well as in terms of computational demands. Evaluations are performed using corpora of Carnatic music from India and a collection of Ballroom dances. We document that the approximate methods perform similar to exact inference, at a lower computational cost. Furthermore, we show that the inference schemes remain accurate for long and full length recordings in Carnatic music.

PS1.28 Predictive Power of Personality on Music-Genre Exclusivity

Jotthi Bansal, Matthew Woolhouse

Studies reveal a strong relationship between personality and preferred musical genre. Our study explored this relationship using a new methodology: genre dispersion among people's mobile-phone music collections. By analyzing the download behaviours of genre-defined user subgroups, we investigated the following questions: (1) do genre-preferring subgroups show distinct patterns of genre consumption and genre exclusivity; (2) does genre exclusivity relate to Big Five personality factors? We hypothesized that genre-preferring subgroups would vary in genre exclusivity, and that their degree of exclusivity would be linearly associated with the openness personality factor (if people have open personalities, they should be "open" to different musical styles). Consistent with our hypothesis, results showed that greater genre inclusivity, i.e. many genres in people's music collections, positively associated with openness and (unexpectedly) agreeableness, suggesting that individuals with high openness and agreeableness have wider musical tastes than those with low openness and agreeableness. Our study corroborated previous research linking genre preference and personality, and revealed, in a novel way, the predictive power of personality on music-consumption.

PS1.29 A Toolkit for Live Annotation of Opera Performance: Experiences Capturing Wagner's *Ring* Cycle

Kevin R. Page, Terhi Nurmikko-Fuller, Carolin Rindfleisch, David M. Weigl, Richard J. Lewis, Laurence Dreyfus, David De Roure

Performance of a musical work potentially provides a rich source of multimedia material for future investigation, both for musicologists' study of reception and perception, and in improvement of computational methods applied to its analysis. This is particularly true of music theatre, where a traditional recording cannot sufficiently capture the ephemeral phenomena unique to each staging. In this paper we introduce a toolkit developed with, and used by, a musicologist throughout a complete multi-day production of Richard Wagner's Der Ring des Nibelungen. The toolkit is centred on a tablet-based score interface through which the scholar makes notes on the scenic setting of the performance as it unfolds, supplemented by a variety of digital data gathered to structure and index the annotations. We report on our experience developing a system suitable for real-time use by the musicologist, structuring the data for reuse and further investigation using semantic web technologies, and of the practical challenges and compromises of fieldwork within a working theatre. Finally we consider the utility of our tooling from both a user perspective and through an initial quantitative investigation of the data gathered.

PS1.30 Selective Acquisition Techniques for Enculturation-Based Melodic Phrase Segmentation

Marcelo E. Rodríguez-López, Anja Volk

Automatic melody segmentation is an important yet unsolved problem in Music Information Retrieval. Research in the field of Music Cognition suggests that previous listening experience plays a considerable role in the perception of melodic segment structure. At present automatic melody segmenters that model listening experience commonly do so using unsupervised statistical learning with 'non-selective' information acquisition techniques, i.e. the learners gather and store information indiscriminately into memory.

In this paper we investigate techniques for 'selective' information acquisition, i.e. our learning model uses a goaloriented approach to select what to store in memory. We test the usefulness of the segmentations produced using selective acquisition learning in a melody classification experiment involving melodies of different cultures. Our results show that the segments produced by our selective learner segmenters substantially improve classification accuracy when compared to segments produced by a nonselective learner segmenter, two local segmentation methods, and two naïve baselines.

Oral Session 1
Corpus Analysis & Annotation

14:30-16:00 Tuesday, October 27

Session Chair: Cynthia C.S. Liem

OS1.1 Corpus Analysis Tools for Computational Hook Discovery

Jan Van Balen, John Ashley Burgoyne, Dimitrios Bountouridis, Daniel Müllensiefen, Remco C. Veltkamp

Compared to studies with symbolic music data, advances in music description from audio have overwhelmingly focused on ground truth reconstruction and maximizing prediction accuracy, with only a small fraction of studies using audio description to gain insight into musical data. We present a strategy for the corpus analysis of audio data that is optimized for interpretable results. The approach brings two previously unexplored concepts to the audio domain: audio bigram distributions, and the use of corpus-relative or "second-order" descriptors. To test the real-world applicability of our method, we present an experiment in which we model song recognition data collected in a widely-played music game. By using the proposed corpus analysis pipeline we are able to present a cognitively adequate analysis that allows a model interpretation in terms of the listening history and experience of our participants. We find that our corpus-based audio features are able to explain a comparable amount of variance to symbolic features for this task when used alone and that they can supplement symbolic features profitably when the two types of features are used in tandem. Finally, we highlight new insights into what makes music recognizable.

OS1.2 Large-Scale Content-Based Matching of MIDI and Audio Files

Colin Raffel, Daniel P. W. Ellis

MIDI files, when paired with corresponding audio recordings, can be used as ground truth for many music information retrieval tasks. We present a system which can efficiently match and align MIDI files to entries in a large corpus of audio content based solely on content, i.e., without using any metadata. The core of our approach is a convolutional network-based cross-modality hashing scheme which transforms feature matrices into sequences of vectors in a common Hamming space. Once represented in this way, we can efficiently perform large-scale dynamic time warping searches to match MIDI data to audio recordings. We evaluate our approach on the task of matching a huge corpus of MIDI files to the Million Song Dataset.

OS1.3 Improving Genre Annotations for the Million Song Dataset

Hendrik Schreibers

Any automatic music genre recognition (MGR) system must show its value in tests against a ground truth dataset. Recently, the public dataset most often used for this purpose has been proven problematic, because of mislabeling, duplications, and its relatively small size. Another dataset, the Million Song Dataset (MSD), a collection of features and metadata for one million tracks, unfortunately does not contain readily accessible genre labels. Therefore, multiple attempts have been made to add song-level genre annotations, which are required for supervised machine learning tasks. Thus far, the quality of these annotations has not been evaluated.

In this paper we present a method for creating additional genre annotations for the MSD from databases, which contain multiple, crowd-sourced genre labels per song (Last.fm, beaTunes). Based on label co-occurrence rates, we derive taxonomies, which allow inference of toplevel genres. These are most often used in MGR systems.

We then combine multiple datasets using majority voting. This both promises a more reliable ground truth and allows the evaluation of the newly generated and preexisting datasets. To facilitate further research, all derived genre annotations are publicly available on our website.

OS1.4 A Software Framework for Musical Data Augmentation

Brian McFee, Eric J. Humphrey, Juan P. Bello

Predictive models for music annotation tasks are practically limited by a paucity of well-annotated training data. In the broader context of large-scale machine learning, the concept of "data augmentation" — supplementing a training set with carefully perturbed samples—has emerged as an important component of robust systems. In this work, we develop a general software framework for augmenting annotated musical datasets, which will allow practitioners to easily expand training sets with musically motivated perturbations of both audio and annotations. As a proof of concept, we investigate the effects of data augmentation on the task of recognizing instruments in mixed signals.

Oral Session 2
Rhythm & Beat

09:00-10:00 Wednesday, October 28

Session Chair: Sebastian Böck

OS2.1 Drum Transcription Using Partially Fixed Non-Negative Matrix Factorization with Template Adaptation

Chih-Wei Wu, Alexander Lerch

In this paper, a template adaptive drum transcription algorithm using partially fixed Non-negative Matrix Factorization (NMF) is presented. The proposed method detects percussive events in complex mixtures of music with a minimal training set. The algorithm decomposes the music signal into two dictionaries: a percussive dictionary initialized with pre-defined drum templates and a harmonic dictionary initialized with undefined entries. The harmonic dictionary is adapted to the non-percussive music content in a standard NMF procedure. The percussive dictionary is adapted to each individual signal in an iterative scheme: it is fixed during the decomposition process, and is updated based on the result of the previous convergence. Two template adaptation methods are proposed to provide more flexibility and robustness in the case of unknown data. The performance of the proposed system has been evaluated and compared to state of the art systems. The results show that template adaptation improves the transcription performance, and the detection accuracy is in the same range as more complex systems.

OS2.2 Beat and Downbeat Tracking Based on Rhythmic Patterns Applied to the Uruguayan Candombe Drumming

Leonardo Nunes, Martín Rocamora, Luis Jure, Luiz W. P. Biscainho

Computational analysis of the rhythmic/metrical structure of music from recorded audio is a hot research topic in music information retrieval. Recent research has explored the explicit modeling of characteristic rhythmic patterns as a way to improve upon existing beat-tracking algorithms, which typically fail on dealing with syncopated or polyrhythmic music. This work takes the Uruguayan Candombe drumming (an afro-rooted rhythm from Latin America) as a case study. After analyzing the aspects that make this music genre troublesome for usual algorithmic approaches and describing its basic rhythmic patterns, the paper proposes a supervised scheme for rhythmic pattern tracking that aims at finding the metric structure from a Candombe recording, including beat and downbeat phases. Then it evaluates and compares the performance of the method with those of general-purpose beat-tracking algorithms through a set of experiments involving a database of annotated recordings totaling over two hours of audio. The results of this work reinforce the advantages of tracking rhythmic patterns (possibly

learned from annotated music) when it comes to automatically following complex rhythms. A software implementation of the proposal as well as the annotated database utilized are available to the research community with the publication of this paper.

OS2.3 Automated Estimation of Ride Cymbal Swing Ratios in Jazz Recordings

Christian Dittmar, Martin Pfleiderer, Meinard Müller

In this paper, we propose a new method suitable for the automatic analysis of microtiming played by drummers in jazz recordings. Specifically, we aim to estimate the drummers' swing ratio in excerpts of jazz recordings taken from the Weimar Jazz Database. A first approach is based on automatic detection of ride cymbal (RC) onsets and evaluation of relative time intervals between them. However, small errors in the onset detection propagate considerably into the swing ratio estimates. As our main technical contribution, we propose to use the log-lag autocorrelation function (LLACF) as a mid-level representation for estimating swing ratios, circumventing the error-prone detection of RC onsets. In our experiments, the LLACF-based swing ratio estimates prove to be more reliable than the ones based on RC onset detection. Therefore, the LLACF seems to be the method of choice to process large amounts of jazz recordings. Finally, we indicate some implications of our method for microtiming studies in jazz research.

Poster Session 2 10:00-11:30, 16:00-17:30 Wednesday, October 28

PS2.1 Musical Offset Detection of Pitched Instruments: The Case of Violin

Che-Yuan Liang, Li Su, Yi-Hsuan Yang, Hsin-Ming Lin

Musical offset detection is an integral part of a music signal processing system that requires complete characterization of note events. However, unlike onset detection, offset detection has seldom been the subject of an in-depth study in the music information retrieval community, possibly because of the ambiguity involved in the determination of offset times in music. This paper presents a preliminary study aiming at discussing ways to annotate and to evaluate offset times for pitched non-percussive instruments. Moreover, we conduct a case study of offset detection in violin recordings by evaluating a number of energy, spectral flux, and pitch based methods using a new dataset covering 6 different violin playing techniques. The new dataset, which is going to be shared with the research community, consists of 63 violin recordings that are thoroughly annotated based on perceptual loudness and note transition. The offset detection methods, which are adapted from wellknown methods for onset detection, are evaluated using an onset-aware method we propose for this task. Result shows that the accuracy of offset detection is highly dependent on the playing techniques involved. Moreover, pitch-based methods can better get rid of the soft-decaying behavior of offsets and achieve the best result among others.

PS2.2 Specter: Combining Music Information Retrieval with Sound Spatialization

Bill Manaris, Seth Stoudenmier

Specter combines music information retrieval (MIR) with sound spatialization to provide a simple, yet versatile environment to experiment with sound spatialization for music composition and live performance. Through various interfaces and sensors, users may position sounds at arbitrary locations and trajectories in a three-dimensional plane. The system utilizes the JythonMusic environment for symbolic music processing, music information retrieval, and live audio manipulation. It also incorporates Iannix, a 3D graphical, open-source sequencer, for realtime generation, manipulation, and storing of sound trajectory scores. Finally, through Glaser, a sound manipulation instrument, Specter renders the various sounds in space. The system architecture supports different sound spatialization techniques including Ambisonics and Vector Based Amplitude Panning. Various interfaces are discussed, including a Kinect-based sensor system, a Leap-Motion-based hand-tracking interface, and a smartphonebased OSC controller. Finally, we present Migrant, a music composition, which utilizes and demon-

strates Specter's ability to combine MIR techniques with sound spatialization through inexpensive, minimal hardware.

PS2.3 Content-Aware Collaborative Music Recommendation Using Pre-Trained Neural Networks

Dawen Liang, Minshu Zhan, Daniel P. W. Ellis

Although content is fundamental to our music listening preferences, the leading performance in music recommendation is achieved by collaborative-filtering-based methods which exploit the similarity patterns in user's listening history rather than the audio content of songs. Meanwhile, collaborative filtering has the well-known "cold-start" problem, i.e., it is unable to work with new songs that no one has listened to. Efforts on incorporating content information into collaborative filtering methods have shown success in many non-musical applications, such as scientific article recommendation. Inspired by the related work, we train a neural network on semantic tagging information as a content model and use it as a prior in a collaborative filtering model. Such a system still allows the user listening data to "speak for itself". The proposed system is evaluated on the Million Song Dataset and shows comparably better result than the collaborative filtering approaches, in addition to the favorable performance in the cold-start case.

PS2.4 Comparative Analysis of Orchestral Performance Recordings: An Image-Based Approach

Cynthia C. S. Liem, Alan Hanjalic

Traditionally, the computer-assisted comparison of multiple performances of the same piece focused on performances on single instruments. Due to data availability, there also has been a strong bias towards analyzing piano performances, in which local timing, dynamics and articulation are important expressive performance features. In this paper, we consider the problem of analyzing multiple performances of the same symphonic piece, performed by different orchestras and different conductors. While differences between interpretations in this genre may include commonly studied features on timing, dynamics and articulation, the timbre of the orchestra and choices of balance within the ensemble are other important aspects distinguishing different orchestral interpretations from one another. While it is hard to model these higher-level aspects as explicit audio features, they can usually be noted visually in spectrogram plots. We therefore propose a method to compare orchestra performances by examining visual spectrogram characteristics. Inspired by eigenfaces in human face recognition, we apply Principal Components Analysis on synchronized performance fragments to localize areas of cross-performance variation in time and frequency. We discuss how this information can be used to examine performer differences, and

how beyond pairwise comparison, relative differences can be studied between multiple performances in a corpus at once.

PS2.5 Monaural Blind Source Separation in the Context of Vocal Detection
Bernhard Lehner, Gerhard Widmer

In this paper, we evaluate the usefulness of several monaural blind source separation (BSS) algorithms in the context of vocal detection (VD). BSS is the problem of recovering several sources, given only a mixture. VD is the problem of automatically identifying the parts in a mixed audio signal, where at least one person is singing. We compare the results of three different strategies for utilising the estimated singing voice signals from four state-of-the-art source separation algorithms. In order to assess the performance of those strategies on an internal data set, we use two different feature sets, each fed to two different classifiers. After selecting the most promising approach, the results on two publicly available data sets are presented. In an additional experiment, we use the improved VD for a simple postprocessing technique: For the final estimation of the source signals, we decide to use either silence, or the mixed, or the separated signals, according to the VD. The results of traditionally used BSS evaluation methods suggest that this is useful for both the estimated background signals, as well as for the estimated vocals.

PS2.6 Detection of Common Mistakes in Novice Violin Playing
Yin-Jyun Luo, Li Su, Yi-Hsuan Yang, Tai-Shih Chi

Analyzing and modeling playing mistakes are essential parts of computer-aided education tools in learning musical instruments. In this paper, we present a system for identifying four types of mistakes commonly made by novice violin players. We construct a new dataset comprising of 981 legato notes played by 10 players across different skill levels, and have violin experts annotate all possible mistakes associated with each note by listening to the recordings. Five feature representations are generated from the same feature set with different scales, including two note-level representations and three segmentlevel representations of the onset, sustain and offset, and are tested for automatically identifying playing mistakes. Performance is evaluated under the framework of using the Fisher score for feature selection and the support vector machine for classification. Results show that the Fmeasures using different feature representations can vary up to 20% for two types of playing mistakes. It demonstrates the different sensitivities of each feature representation to different mistakes. Moreover, our results suggest that the standard audio features such as MFCCs are not good enough and more advanced feature design may be needed.

PS2.7 Probabilistic Modular Bass Voice Leading in Melodic Harmonisation

Dimos Makris, Maximos Kaliakatsos-Papakostas, Emilios Cambouropoulos

Probabilistic methodologies provide successful tools for automated music composition, such as melodic harmonisation, since they capture statistical rules of the music idioms they are trained with. Proposed methodologies focus either on specific aspects of harmony (e.g., generating abstract chord symbols) or incorporate the determination of many harmonic characteristics in a single probabilistic generative scheme. This paper addresses the problem of assigning voice leading focussing on the bass voice, i.e., the realisation of the actual bass pitches of an abstract chord sequence, under the scope of a modular melodic harmonisation system where different aspects of the generative process are arranged by different modules. The proposed technique defines the motion of the bass voice according to several statistical aspects: melody voice contour, previous bass line motion, bass-to-melody distances and statistics regarding inversions and note doublings in chords. The aforementioned aspects of voicing are modular, i.e., each criterion is defined by independent statistical learning tools. Experimental results on diverse music idioms indicate that the proposed methodology captures efficiently the voice layout characteristics of each idiom, whilst additional analyses on separate statistically trained modules reveal distinctive aspects of each idiom. The proposed system is designed to be flexible and adaptable (for instance, for the generation of novel blended melodic harmonisations).

PS2.8 An Iterative Multi Range Non-Negative Matrix Factorization Algorithm for Polyphonic Music Transcription

Anis Khlif, Vidhyasaharan Sethu

This article presents a novel iterative algorithm based on Non-negative Matrix Factorisation (NMF) that is particularly well suited to the task of automatic music transcription (AMT). Compared with previous NMF based techniques, this one does not aim at factorizing the timefrequency representation of the entire musical signal into a combination of the possible set of notes. Instead, the proposed algorithm proceeds iteratively by initially decomposing a part of the time-frequency representation into a combination of a small subset of all possible notes then reinvesting this information in the following step involving a large subset of notes. Specifically, starting with the lowest octave of notes that is of interest, each iteration increases the set of notes under consideration by an octave. The resolution of a lower dimensionality problem used to properly initialize matrices for a more complex problem, results in a gain of some percent in the transcription accuracy.

PS2.9 Training Phoneme Models for Singing with "Songified" Speech Data
Anna M. Kruspe

Speech recognition in singing is a task that has not been widely researched so far. Singing possesses several characteristics that differentiate it from speech. Therefore, algorithms and models that were developed for speech usually perform worse on singing. One of the bottlenecks in many algorithms is the recognition of phonemes in singing. We noticed that this recognition step can be improved when using singing data in model training, but to our knowledge, there are no large datasets of singing data annotated with phonemes. However, such data does exist for speech. We therefore propose to make phoneme recognition models more robust for singing by training them on speech data that has artificially been made more "song-like". We test two main modifications on speech data: Time stretching and pitch shifting. Artificial vibrato is also tested. We then evaluate models trained on different combinations of these modified speech recordings. The utilized modeling algorithms are Neural Networks and Deep Belief Networks.

PS2.10 Graph-Based Rhythm Interpretation
Rong Jin, Christopher Raphael

We present a system that interprets the notated rhythm obtained from optical music recognition (OMR). Our approach represents the notes and rests in a system measure as the vertices of a graph. We connect the graph by adding voice edges and coincidence edges between pairs of vertices, while the rhythmic interpretation follows simply from the connected graph. The graph identification problem is cast as an optimization where each potential edge is scored according to its plausibility. We seek the optimally scoring graph where the score is represented as a sum of edge scores. Experiments were performed on about 60 score pages showing that our system can handle difficult rhythmic situations including multiple voices, voices that merge and split, voices spanning two staves, and missing tuplets.

PS2.11 Let it Bee – Towards NMF-Inspired Audio Mosaicing
Jonathan Driedger, Thomas Prätzlich, Meinard Müller

A swarm of bees buzzing "Let it be" by the Beatles or the wind gently howling the romantic "Gute Nacht" by Schubert – these are examples of *audio mosaics* as we want to create them. Given a *target* and a *source* recording, the goal of audio mosaicing is to generate a *mosaic* recording that conveys musical aspects (like melody and rhythm) of the target, using sound components taken from the source. In this work, we propose a novel approach for automatically generating audio mosaics with the objective to preserve the source's timbre in the mosaic.

Inspired by algorithms for *non-negative matrix factorization* (NMF), our idea is to use update rules to learn an activation matrix that, when multiplied with the spectrogram of the source recording, resembles the spectrogram of the target recording. However, when applying the original NMF procedure, the resulting mosaic does not adequately reflect the source's timbre. As our main technical contribution, we propose an extended set of update rules for the iterative learning procedure that supports the development of sparse diagonal structures in the activation matrix. We show how these structures better retain the source's timbral characteristics in the resulting mosaic.

PS2.12 Real-Time Music Tracking Using Multiple Performances as a Reference

Andreas Arzt, Gerhard Widmer

In general, algorithms for real-time music tracking directly use a symbolic representation of the score, or a synthesised version thereof, as a reference for the on-line alignment process. In this paper we present an alternative approach. First, different performances of the piece in question are collected and aligned (off-line) to the symbolic score. Then, multiple instances of the on-line tracking algorithm (each using a different performance as a reference) are used to follow the live performance, and their output is combined to come up with the current position in the score. As the evaluation shows, this strategy improves both the robustness and the precision, especially on pieces that are generally hard to track (e.g. pieces with extreme, abrupt tempo changes, or orchestral pieces with a high degree of polyphony). Finally, we describe a real-world application, where this music tracking algorithm was used to follow a world-famous orchestra in a concert hall in order to show synchronised visual content (the sheet music, explanatory text and videos) to members of the audience.

PS2.13 Two Data Sets for Tempo Estimation and Key Detection in Electronic Dance Music Annotated from User Corrections

Peter Knees, Ángel Faraldo, Perfecto Herrera, Richard Vogl, Sebastian Böck, Florian Hörschläger, Mickael Le Goff

We present two new data sets for automatic evaluation of tempo estimation and key detection algorithms. In contrast to existing collections, both released data sets focus on electronic dance music (EDM). The data sets have been automatically created from user feedback and annotations extracted from web sources. More precisely, we utilize user corrections submitted to an online forum to report wrong tempo and key annotations on the Beatport website. Beatport is a digital record store targeted at DJs and focusing on EDM genres. For all annotated tracks in the data sets, samples of at least one-minute-length can be freely downloaded. For key detection, further ground truth is extracted from expert an-

notations manually assigned to Beatport tracks for benchmarking purposes. The set for tempo estimation comprises 664 tracks and the set for key detection 604 tracks. We detail the creation process of both data sets and perform extensive benchmarks using state-of-theart algorithms from both academic research and commercial products.

PS2.14 Towards Support for Understanding Classical Music: Alignment of Content Descriptions on the Web

Taku Kuribayashi, Yasuhito Asano, Masatoshi Yoshikawa

Supporting the understanding of classical music is an important topic that involves various research fields such as text analysis and acoustics analysis. Content descriptions are explanations of classical music compositions that help a person to understand technical aspects of the music. Recently, Kuribayashi et al. proposed a method for obtaining content descriptions from the web. However, the content descriptions on a single page frequently explain a specific part of a composition only. Therefore, a person who wants to fully understand the composition suffers from a time-consuming task, which seems almost impossible for a novice of classical music. To integrate the content descriptions obtained from multiple pages, we propose a method for aligning each pair of paragraphs of such descriptions. Using dynamic time warping-based method along with our new ideas, (a) a distribution-based distance measure named w2DD, and (b) the concept of passage expressions, it is possible to align content descriptions of classical music better than when using cutting-edge text analysis methods. Our method can be extended in future studies to create applications systems to integrate descriptions with musical scores and performances.

PS2.15 FlaBase: Towards the Creation of a Flamenco Music Knowledge Base

Sergio Oramas, Francisco Gómez, Emilia Gómez, Joaquín Mora

Online information about flamenco music is scattered over different sites and knowledge bases. Unfortunately, there is no common repository that indexes all these data. In this work, information related to flamenco music is gathered from general knowledge bases (e.g., Wikipedia, DBpedia), music encyclopedias (e.g., MusicBrainz), and specialized flamenco websites, and is then integrated into a new knowledge base called FlaBase. As resources from different data sources do not share common identifiers, a process of pair-wise entity resolution has been performed. FlaBase contains information about 1,174 artists, 76 palos (flamenco genres), 2,913 albums, 14,078 tracks, and 771 Andalusian locations. It is freely available in RDF and JSON formats. In addition, a method for entity recognition and disambiguation for FlaBase has been created. The system can recognize and disambiguate FlaBase entity references in Spanish texts with an f-

measure value of 0.77. We applied it to biographical texts present in Flabase. By using the extracted information, the knowledge base is populated with relevant information and a semantic graph is created connecting the entities of FlaBase. Artists relevance is then computed over the graph and evaluated according to a flamenco expert criteria. Accuracy of results shows a high degree of quality and completeness of the knowledge base.

PS2.16 Discovery of Syllabic Percussion Patterns in Tabla Solo Recordings
Swapnil Gupta, Ajay Srinivasamurthy, Manoj Kumar, Hema A. Murthy, Xavier Serra

We address the unexplored problem of percussion pattern discovery in Indian art music. Percussion in Indian art music uses onomatopoeic oral mnemonic syllables for the transmission of repertoire and technique. This is utilized for the task of percussion pattern discovery from audio recordings. From a parallel corpus of audio and expert curated scores for 38 tabla solo recordings, we use the scores to build a set of most frequent syllabic patterns of different lengths. From this set, we manually select a subset of musically representative query patterns. To discover these query patterns in an audio recording, we use syllable-level hidden Markov models (HMM) to automatically transcribe the recording into a syllable sequence, in which we search for the query pattern instances using a Rough Longest Common Subsequence (RLCS) approach. We show that the use of RLCS makes the approach robust to errors in automatic transcription, significantly improving the pattern recall rate and Fmeasure. We further propose possible enhancements to improve the results.

PS2.17 Autoregressive Hidden Semi-Markov Model of Symbolic Music Performance for Score Following
Eita Nakamura, Philippe Cuvillier, Arshia Cont, Nobutaka Ono, Shigeki Sagayama

A stochastic model of symbolic (MIDI) performance of polyphonic scores is presented and applied to score following. Stochastic modelling has been one of the most successful strategies in this field. We describe the performance as a hierarchical process of performer's progression in the score and the production of performed notes, and represent the process as an extension of the hidden semi-Markov model. The model is compared with a previously studied model based on hidden Markov model (HMM), and reasons are given that the present model is advantageous for score following especially for scores with trills, tremolos, and arpeggios. This is also confirmed empirically by comparing the accuracy of score following and analysing the errors. We also provide a hybrid of this model and the HMM-based model which is computationally more efficient and retains the advantages of the former model. The present model yields one of the state-

of-the-art score following algorithms for symbolic performance and can possibly be applicable for other music recognition problems.

PS2.18 Automatic Mashup Creation by Considering both Vertical and Horizontal Mashabilities

Chuan-Lung Lee, Yin-Tzu Lin, Zun-Ren Yao, Feng-Yi Lee, Ja-Ling Wu

In this paper, we proposed a system to effectively create music mashups – a kind of re-created music that is made by mixing parts of multiple existing music pieces. Unlike previous studies which merely generate mashups by overlaying music segments on one single base track, the proposed system creates mashups with multiple background (e.g. instrumental) and lead (e.g. vocal) track segments. So, besides the suitability between the vertically overlaid tracks (i.e. vertical mashability) used in previous studies, we proposed to further consider the suitability between the horizontally connected consecutive music segments (i.e. horizontal mashability) when searching for proper music segments to be combined. On the vertical side, two new factors: "harmonic change balance" and "volume weight" have been considered. On the horizontal side, the methods used in the studies of medley creation are incorporated. Combining vertical and horizontal mashabilities together, we defined four levels of mashability that may be encountered and found the proper solution to each of them. Subjective evaluations showed that the proposed four levels of mashability can appropriately reflect the degrees of listening enjoyment. Besides, by taking the newly proposed vertical mashability measurement into account, the improvement in user satisfaction is statistically significant.

PS2.19 Hierarchical Evaluation of Segment Boundary Detection

Brian McFee, Oriol Nieto, Juan P. Bello

Structure in music is traditionally analyzed hierarchically: large-scale sections can be sub-divided and refined down to the short melodic ideas at the motivic level. However, typical algorithmic approaches to structural annotation produce flat temporal partitions of a track, which are commonly evaluated against a similarly flat, human-produced annotation. Evaluating structure analysis as represented by flat annotations effectively discards all notions of structural depth in the evaluation. Although collections of hierarchical structure annotations have been recently published, no techniques yet exist to measure an algorithm's accuracy against these rich structural annotations. In this work, we propose a method to evaluate structural boundary detection with hierarchical annotations. The proposed method transforms boundary detection into a ranking problem, and facilitates the comparison of both flat and hierarchical annotations. We demonstrate

the behavior of the proposed method with various synthetic and real examples drawn from the SALAMI dataset.

PS2.20 Improving MIDI Guitar's Accuracy with NMF and Neural Net
Masaki Otsuka, Tetsuro Kitahara

In this paper, we propose a method for improving the accuracy of MIDI guitars. MIDI guitars are useful tools for various purposes from inputting MIDI data to enjoying a jam session system, but existing MIDI guitars do not have sufficient accuracy in converting the performance to an MIDI form. In this paper, we make an attempt on improving the accuracy of a MIDI guitar by integrating it with an audio transcription method based on non-negative matrix factorization (NMF). First, we investigate an NMF-based algorithm for transcribing guitar performances. Although the NMF is a promising method, an effective post-process (i.e., converting the NMF's output to an MIDI form) is a non-trivial problem. We propose use of a neural network for this conversion. Next, we investigate a method for integrating the outputs of the MIDI guitar and NMF. Because they have different tendencies in wrong outputs, we take an policy of outputting only common parts in the two outputs. Experimental results showed that the F-score of our method was 0.626 whereas those of the MIDI-guitar-only and NMF-and-neural-network-only methods were 0.347 and 0.526, respectively.

PS2.21 Analysis of Intonation Trajectories in Solo Singing
Jiajie Dai, Matthias Mauch, Simon Dixon

We present a new dataset for singing analysis and modelling, and an exploratory analysis of pitch accuracy and pitch trajectories. Shortened versions of three pieces from The Sound of Music were selected: "Edelweiss", "Do-Re- Mi" and "My Favourite Things". 39 participants sang three repetitions of each excerpt without accompaniment, resulting in a dataset of 21762 notes in 117 recordings. To obtain pitch estimates we used the Tony software's automatic transcription and manual correction tools. Pitch accuracy was measured in terms of pitch error and interval error. We show that singers' pitch accuracy correlates significantly with self-reported singing skill and musical training. Larger intervals led to larger errors, and the tritone interval in particular led to average errors of one third of a semitone. Note duration (or inter-onset interval) had a significant effect on pitch accuracy, with greater accuracy on longer notes. To model drift in the tonal centre over time, we present a sliding window model which reveals patterns in the pitch errors of some singers. Based on the trajectory, we propose a measure for the magnitude of drift: tonal reference deviation (TRD). The data and software are freely available.

PS2.22 Evaluating the General Chord Type Representation in Tonal Music and Organising GCT Chord Labels in Functional Chord Categories

Maximos Kaliakatsos-Papakostas, Asterios Zacharakis, Costas Tsougras, Emilios Cambouropoulos

The General Chord Type (GCT) representation is appropriate for encoding tone simultaneities in any harmonic context (such as tonal, modal, jazz, octatonic, atonal). The GCT allows the re-arrangement of the notes of a harmonic sonority such that abstract idiom-specific types of chords may be derived. This encoding is inspired by the standard roman numeral chord type labelling and is, therefore, ideal for hierarchic harmonic systems such as the tonal system and its many variations; at the same time, it adjusts to any other harmonic system such as post-tonal, atonal music, or traditional polyphonic systems. In this paper the descriptive potential of the GCT is assessed in the tonal idiom by comparing GCT harmonic labels with human expert annotations (Kostka & Payne harmonic dataset). Additionally, novel methods for grouping and clustering chords, according to their GCT encoding and their functional role in chord sequences, are introduced. The results of both harmonic labelling and functional clustering indicate that the GCT representation constitutes a suitable scheme for representing effectively harmony in computational systems.

PS2.23 Beat Histogram Features from NMF-Based Novelty Functions for Music Classification

Athanasios Lykartsis, Chih-Wei Wu, Alexander Lerch

In this paper we present novel rhythm features derived from drum tracks extracted from polyphonic music and evaluate them in a genre classification task. Musical excerpts are analyzed using an optimized, partially fixed Non-Negative Matrix Factorization (NMF) method and beat histogram features are calculated on basis of the resulting activation functions for each one out of three drum tracks extracted (Hi-Hat, Snare Drum and Bass Drum). The features are evaluated on two widely used genre datasets (GTZAN and Ballroom) using standard classification methods, concerning the achieved overall classification accuracy. Furthermore, their suitability in distinguishing between rhythmically similar genres and the performance of the features resulting from individual activation functions is discussed. Results show that the presented NMF-based beat histogram features can provide comparable performance to other classification systems, while considering strictly drum patterns.

PS2.24 Music Shapelets for Fast Cover Song Recognition

Diego F. Silva, Vinícius M. A. Souza, Gustavo E. A. P. A. Batista

A cover song is a new performance or recording of a previously recorded music by an artist other than the original one. The automatic identification of cover songs is useful for a wide range of tasks, from fans looking for new versions of their favorite songs to organizations involved in licensing copyrighted songs. This is a difficult task given that a cover may differ from the original song in key, timbre, tempo, structure, arrangement and even language of the vocals. Cover song identification has attracted some attention recently. However, most of the state-of-the-art approaches are based on similarity search, which involves a large number of similarity computations to retrieve potential cover versions for a query recording. In this paper, we adapt the idea of time series shapelets for content-based music retrieval. Our proposal adds a training phase that finds small excerpts of feature vectors that best describe each song. We demonstrate that we can use such small segments to identify cover songs with higher identification rates and more than one order of magnitude faster than methods that use features to describe the whole music.

PS2.25 Improving Score-Informed Source Separation for Classical Music Through Note Refinement

Marius Miron, Julio José Carabias-Orti, Jordi Janer

Signal decomposition methods such as Non-negative Matrix Factorization (NMF) demonstrated to be a suitable approach for music signal processing applications, including sound source separation. To better control this decomposition, NMF has been extended using prior knowledge and parametric models. In fact, using score information considerably improved separation results. Nevertheless, one of the main problems of using score information is the misalignment between the score and the actual performance. A potential solution to this problem is the use of audio to score alignment systems. However, most of them rely on a tolerance window that clearly affects the separation results. To overcome this problem, we propose a novel method to refine the aligned score at note level by detecting both, onset and offset for each note present in the score. Note refinement is achieved by detecting shapes and contours in the estimated instrument-wise time activation (gains) matrix. Decomposition is performed in a supervised way, using training instrument models and coarsely-aligned score information. The detected contours define time-frequency note boundaries, and they increase the sparsity. Finally, we have evaluated our method for informed source separation using a dataset of Bach chorales obtaining satisfactory results, especially in terms of SIR.

PS2.26 In their Own Words: Using Text Analysis to Identify Musicologists' Attitudes towards Technology

Charles Inskip, Frans Wiering

A widely distributed online survey gathered quantitative and qualitative data relating to the use of technology in the research practices of musicologists. This survey builds on existing work in the digital humanities and provides insights into the specific nature of musicology in relation to use and perceptions of technology. Analysis of the data (n=621) notes the preferences in resource format and the digital skills of the survey participants. The themes of comments on rewards, benefits, frustrations, risks, and limitations are explored using an h-point approach derived from applied linguistics. It is suggested that the research practices of musicologists reflect wider existing research into the digital humanities, and that efforts should be made into supporting development of their digital skills and providing usable, useful and reliable software created with a 'musicology-centred' design approach. This software should support online access to high quality digital resources (image, text, sound) which are comprehensive and discoverable, and can be shared, reused and manipulated at a micro- and macro level.

PS2.27 Combining Features for Cover Song Identification

Julien Osmalskyj, Peter Foster, Simon Dixon, Jean-Jacques Embrechts

In this paper, we evaluate a set of methods for combining features for cover song identification. We first create multiple classifiers based on global tempo, duration, loudness, beats and chroma average features, training a random forest for each feature. Subsequently, we evaluate standard combination rules for merging these single classifiers into a composite classifier based on global features. We further obtain two higher level classifiers based on chroma features: one based on comparing histograms of quantized chroma features, and a second one based on computing cross-correlations between sequences of chroma features, to account for temporal information. For combining the latter chroma-based classifiers with the composite classifier based on global features, we use standard rank aggregation methods adapted from the information retrieval literature. We evaluate performance with the Second Hand Song dataset, where we quantify performance using multiple statistics. We observe that each combination rule outperforms single methods in terms of the total number of identified queries. Experiments with rank aggregation methods show an increase of up to 23.5 % of the number of identified queries, compared to single classifiers.

PS2.28 Score Following for Piano Performances with Sustain-Pedal Effects
Bochen Li, Zhiyao Duan

One challenge in score following (i.e., mapping audio frames to score positions in real time) for piano performances is the mismatch between audio and score caused by the usage of the sustain pedal. When the pedal is pressed, notes played will continue to sound until the string vibration naturally ceases. This makes the notes longer than their notated lengths and overlap with later notes. In this paper, we propose an approach to address this problem. Given that the most competitive wrong score positions for each audio frame are the ones before the correct position due to the sustained sounds, we remove partials of sustained notes and only retain partials of "new notes" in the audio representation. This operation reduces sustain-pedal effects by weakening the match between the audio frame and previous wrong score positions, hence encourages the system to align to the correct score position. We implement this idea based on a state-of-the-art score following framework. Experiments on synthetic and real piano performances from the MAPS dataset show significant improvements on both alignment accuracy and robustness.

PS2.29 Understanding Users of Commercial Music Services through Personas: Design Implications
Jin Ha Lee, Rachel Price

Most of the previous literature on music users' needs, habits, and interactions with music information retrieval (MIR) systems focuses on investigating user groups of particular demographics or testing the usability of specific interfaces/systems. In order to improve our understanding of how users' personalities and characteristics affect their needs and interactions with MIR systems, we conducted a qualitative user study across multiple commercial music services, utilizing interviews and think-aloud sessions. Based on the empirical user data, we have developed seven personas. These personas offer a deeper understanding of the different types of MIR system users and the relative importance of various design implications for each user type. Implications for system design include a renegotiation of our understanding of desired user engagement, especially with the habit of context-switching, designing systems for specialized uses, and addressing user concerns around privacy, transparency, and control.

PS2.30 Corpus-Based Rhythmic Pattern Analysis of Ragtime Syncopation
Hendrik Vincent Koops, Anja Volk, W. Bas de Haas

This paper presents a corpus-based study on rhythmic patterns in the RAG-collection of approximately 11.000 symbolically encoded ragtime pieces. While characteristic musical features that define ragtime as a genre have been debated

since its inception, musicologists argue that specific syncopation patterns are most typical for this genre. Therefore, we investigate the use of syncopation patterns in the RAG-collection from its beginnings until the present time in this paper. Using computational methods, this paper provides an overview on the use of rhythmical patterns of the ragtime genre, thereby offering valuable new insights that complement musicological hypotheses about this genre. Specifically, we measure the amount of syncopation for each bar using Longuet-Higgins and Lee's model of syncopation, determine the most frequent rhythmic patterns, and discuss the role of a specific short-long-short syncopation pattern that musicologists argue is characteristic for ragtime. A comparison between the ragtime (pre-1920) and modern (post-1920) era shows that the two eras differ in syncopation pattern use. Onset density and amount of syncopation increase after 1920. Moreover, our study confirms the musicological hypothesis on the important role of the short-long-short syncopation pattern in ragtime. These findings are pivotal in developing ragtime genre-specific features.

Oral Session 3
Melody & Voice

11:30-12:30 Wednesday, October 28

Session Chair: Matthias Mauch

OS3.1 Comparing Voice and Stream Segmentation Algorithms

Nicolas Guiomard-Kagan, Mathieu Giraud, Richard Groult, Florence Levé

Voice and stream segmentation algorithms group notes from polyphonic data into relevant units, providing a better understanding of a musical score. Voice segmentation algorithms usually extract voices from the beginning to the end of the piece, whereas stream segmentation algorithms identify smaller segments. In both cases, the goal can be to obtain mostly monophonic units, but streams with polyphonic data are also relevant. These algorithms usually cluster contiguous notes with close pitches. We propose an independent evaluation of four of these algorithms (Temperley, Chew and Wu, Ishigaki et al., and Rafailidis et al.) using several evaluation metrics. We benchmark the algorithms on a corpus containing the 48 fugues of Well- Tempered Clavier by J. S. Bach as well as 97 files of popular music containing actual polyphonic information. We discuss how to compare together voice and stream segmentation algorithms, and discuss their strengths and weaknesses.

OS3.2 Melody Extraction by Contour Classification

Rachel M. Bittner, Justin Salamon, Slim Essid, Juan P. Bello

Due to the scarcity of labeled data, most melody extraction algorithms do not rely on fully data-driven processing blocks but rather on careful engineering. For example, the Melodia melody extraction algorithm employs a pitch contour selection stage that relies on a number of heuristics for selecting the melodic output. In this paper we explore the use of a discriminative model to perform purely datadriven melodic contour selection. Specifically, a discriminative binary classifier is trained to distinguish melodic from non-melodic contours. This classifier is then used to predict likelihoods for a track's extracted contours, and these scores are decoded to generate a single melody output. The results are compared with the Melodia algorithm and with a generative model used in a previous study. We show that the discriminative model outperforms the generative model in terms of contour classification accuracy, and the melody output from our proposed system performs comparatively to Melodia. The results are complemented with error analysis and avenues for future improvements.

OS3.3 Comparison of the Singing Style of Two Jingju Schools

Rafael Caro Repetto, Rong Gong, Nadine Kroher, Xavier Serra

Performing schools (*liupai*) in jingju (also known as Peking or Beijing opera) are one of the most important elements for the appreciation of this genre among connoisseurs. In the current paper, we study the potential of MIR techniques for supporting and enhancing musicological descriptions of the singing style of two of the most renowned jingju schools for the *dan* role-type, namely Mei and Cheng schools. To this aim, from the characteristics commonly used for describing singing style in musicological literature, we have selected those that can be studied using standard audio features. We have selected eight recordings from our jingju music research corpus and have applied current algorithms for the measurement of the selected features. Obtained results support the descriptions from musicological sources in all cases but one, and also add precision to them by providing specific measurements. Besides, our methodology suggests some characteristics not accounted for in our musicological sources. Finally, we discuss the need for engaging jingju experts in our future research and applying this approach for musicological and educational purposes as a way of better validating our methodology.

Oral Session 4
Mixed

14:30-16:00 Wednesday, October 28

Session Chair: Eric J. Humphrey

OS4.1 Improving Optical Music Recognition by Combining Outputs from Multiple Sources

Victor Padilla, Alex McLean, Alan Marsden, Kia Ng

Current software for Optical Music Recognition (OMR) produces outputs with too many errors that render it an unrealistic option for the production of a large corpus of symbolic music files. In this paper, we propose a system which applies image pre-processing techniques to scans of scores and combines the outputs of different commercial OMR programs when applied to images of different scores of the same piece of music. As a result of this procedure, the combined output has around 50when compared to the output of any one OMR program. Image pre-processing splits scores into separate movements and sections and removes ossia staves which confuse OMR software. Post-processing aligns the outputs from different OMR programs and from different sources, rejecting outputs with the most errors and using majority voting to determine the likely correct details. Our software produces output in MusicXML, concentrating on accurate pitch and rhythm and ignoring grace notes. Results of tests on the six string quartets by Mozart dedicated to Joseph Haydn and the first six piano sonatas by Mozart are presented, showing an average recognition rate of around 95%.

OS4.2 Relating Natural Language Text to Musical Passages

Richard Sutcliffe, Tim Crawford, Chris Fox, Deane L. Root, Eduard Hovy, Richard J. Lewis

There is a vast body of musicological literature containing detailed analyses of musical works. These texts make frequent references to musical passages in scores by means of natural language phrases. Our longterm aim is to investigate whether these phrases can be linked automatically to the musical passages to which they refer. As a first step, we have organised for two years running a shared evaluation in which participants must develop software to identify passages in a MusicXML score based on a short noun phrase in English. In this paper, we present the rationale for this work, discuss the kind of references to musical passages which can occur in actual scholarly texts, describe the first two years of the evaluation and finally appraise the results to establish what progress we have made.

OS4.3 Music Boundary Detection Using Neural Networks on Combined Features and Two-Level Annotations

Thomas Grill, Jan Schlüter

The determination of structural boundaries is a key task for understanding the structure of a musical piece, but it is also highly ambiguous. Recently, Convolutional Neural Networks (CNN) trained on spectrogram features and human annotations have been successfully used to tackle the problem, but still fall clearly behind human performance. We expand on the CNN approach by combining spectrograms with self-similarity lag matrices as audio features, thereby capturing more facets of the underlying structural information. Furthermore, in order to consider the hierarchical nature of structural organization, we explore different strategies to learn from the two-level annotations of main and secondary boundaries available in the SALAMI structural annotation dataset. We show that both measures improve boundary recognition performance, resulting in a significant improvement over the previous state of the art. As a side-effect, our algorithm can predict boundaries on two different structural levels, equivalent to the training data.

OS4.4 Neuroimaging Methods for Music Information Retrieval: Current Findings and Future Prospects

Blair Kaneshiro, Jacek P. Dmochowski

Over the past decade and a half, music information retrieval (MIR) has grown into a robust, cross-disciplinary field spanning a variety of research domains. Collaborations between MIR and neuroscience researchers, however, are still rare, and to date only a few studies using approaches from one domain have successfully reached an audience in the other. In this paper, we take an initial step toward bridging these two fields by reviewing studies from the music neuroscience literature, with an emphasis on imaging modalities and analysis techniques that might be of practical interest to the MIR community. We show that certain approaches currently used in a neuroscientific setting align with those used in MIR research, and discuss implications for potential areas of future research. We additionally consider the impact of disparate research objectives between the two fields, and how such a discrepancy may have hindered cross-discipline output thus far. It is hoped that a heightened awareness of this literature will foster interaction and collaboration between MIR and neuroscience researchers, leading to advances in both fields that would not have been achieved independently.

Oral Session 5
Similarity

09:00-10:00 Thursday, October 29

Session Chair: Bob Sturm

OS5.1 Improving Visualization of High-Dimensional Music Similarity Spaces

Arthur Flexer

Visualizations of music databases are a popular form of interface allowing intuitive exploration of music catalogs. They are often based on lower dimensional projections of high dimensional music similarity spaces. Such similarity spaces have already been shown to be negatively impacted by so-called hubs and anti-hubs. These are points that appear very close or very far to many other data points due to a problem of measuring distances in high-dimensional spaces. We present an empirical study on how this phenomenon impacts three popular approaches to compute twodimensional visualizations of music databases. We also show how the negative impact of hubs and anti-hubs can be reduced by re-scaling the high dimensional spaces before low dimensional projection.

OS5.2 I-Vectors for Timbre-Based Music Similarity and Music Artist Classification

Hamid Eghbal-zadeh, Bernhard Lehner, Markus Schedl, Gerhard Widmer

In this paper, we present a novel approach to extract songlevel descriptors built from frame-level timbral features such as Mel-frequency cepstral coefficient (MFCC). These descriptors are called identity vectors or i-vectors and are the results of a factor analysis procedure applied on framelevel features. The i-vectors provide a low-dimensional and fixed-length representation for each song and can be used in a supervised and unsupervised manner. First, we use the i-vectors for an unsupervised music similarity estimation, where we calculate the distance between i-vectors in order to predict the genre of songs. Second, for a supervised artist classification task we report the performance measures using multiple classifiers trained on the i-vectors. Standard datasets for each task are used to evaluate our method and the results are compared with the state of the art. By only using timbral information, we already achieved the state of the art performance in music similarity (which uses extra information such as rhythm). In artist classification using timbre descriptors, our method outperformed the state of the art.

OS5.3 Correlating Extracted and Ground-Truth Harmonic Data in Music Retrieval Tasks

Dylan Freedman, Eddie Kohler, Hans Tutschku

We show that traditional music information retrieval tasks with well-chosen parameters perform similarly using computationally extracted chord annotations and groundtruth annotations. Using a collection of Billboard songs with provided ground-truth chord labels, we use established chord identification algorithms to produce a corresponding extracted chord label dataset. We implement methods to compare chord progressions between two songs on the basis of their optimal local alignment scores. We create a set of chord progression comparison parameters defined by chord distance metrics, gap costs, and normalization measures and run a black-box global optimization algorithm to stochastically search for the best parameter set to maximize the rank correlation for two harmonic retrieval tasks across the ground-truth and extracted chord Billboard datasets. The first task evaluates chord progression similarity between all pairwise combinations of songs, separately ranks results for ground-truth and extracted chord labels, and returns a rank correlation coefficient. The second task queries the set of songs with fabricated chord progressions, ranks each query's results across ground-truth and extracted chord labels, and returns rank correlations. The end results suggest that practical retrieval systems can be constructed to work effectively without the guide of human ground-truthing.

Poster Session 3

10:00-11:30, 16:00-17:30 Thursday, October 29

PS3.1 Classical Music on the Web – User Interfaces and Data Representations

Martin Gasser, Andreas Arzt, Thassilo Gadermaier, Maarten Grachten, Gerhard Widmer

We present a set of web-based user interfaces for explorative analysis and visualization of classical orchestral music and a web API that serves as a backend to those applications; we describe use cases that motivated our developments within the PHENICX project, which promotes a vital interaction between Music Information Retrieval research groups and a world-renowned symphony orchestra. Furthermore, we describe two real-world applications that involve the work presented here. Firstly, our web applications are used in the editorial stage of a periodically released subscription-based mobile app by the Royal Concertgebouw Orchestra (RCO) 1 , which serves as a contentdistribution channel for multi-modally enhanced recordings of classical concerts. Secondly, our web API and user interfaces have been successfully used to provide real-time information (such as the score, and explanatory comments from musicologists) to the audience during a live concert of the RCO.

PS3.2 A Statistical View on the Expressive Timing of Piano Rolled Chords

Mutian Fu, Guangyu Xia, Roger B. Dannenberg, Larry Wasserman

Rolled or *arpeggiated* chords are notated chords performed by playing the notes sequentially, usually from lowest to highest in pitch. Arpeggiation is a characteristic of musical expression, or expressive timing, in piano performance. However, very few studies have investigated rolled chord performance. In this paper, we investigate two expressive timing properties of piano rolled chords: *equivalent onset* and *onset span*. Equivalent onset refers to the hidden onset that can functionally replace the onsets of the notes in a chord; onset span refers to the time interval from the first note onset to the last note onset. We ask two research questions. First, what is the equivalent onset of a rolled chord? Second, are the onset spans of different chords interpreted in the same way? The first question is answered by local tempo estimation while the second question is answered by *Analysis of Variance*. Also, we contribute a piano duet dataset for rolled chords analysis and other studies on expressive music performance. The dataset contains three pieces of music, each performed multiple times by different pairs of musicians.

PS3.3 Hybrid Long- and Short-Term Models of Folk Melodies

Srikanth Cherla, Son N. Tran, Tillman Weyde, Artur d'Avila Garcez

In this paper, we present the results of a study on dynamic models for predicting sequences of musical pitch in melodies. Such models predict a probability distribution over the possible values of the next pitch in a sequence, which is obtained by combining the prediction of two components (1) a long-term model (LTM) learned offline on a corpus of melodies, as well as (2) a short-term model (STM) which incorporates context-specific information available during prediction. Both the LTM and the STM learn regularities in pitch sequences solely from data. The models are combined in an ensemble, wherein they are weighted by the relative entropies of their respective predictions. Going by previous work that demonstrates the success of Connectionist LTMs, we employ the recently proposed Recurrent Temporal Discriminative Restricted Boltzmann Machine (RTDRBM) as the LTM here. While it is indeed possible for the same model to also serve as an STM, our experiments showed that n-gram models tended to learn faster than the RTDRBM in an online setting and that the hybrid of an RTDRBM LTM and an n-gram STM gives us the best predictive performance yet on a corpus of monophonic chorale and folk melodies.

PS3.4 Efficient Melodic Query Based Audio Search for Hindustani Vocal Compositions

Kaustuv Kanti Ganguli, Abhinav Rastog, Vedhas Pandit, Prithvi Kantan, Preeti Rao

Time-series pattern matching methods that incorporate time warping have recently been used with varying degrees of success on tasks of search and discovery of melodic phrases from audio for Indian classical vocal music. While these methods perform effectively due to the minimal assumptions they place on the nature of the sampled pitch temporal trajectories, their practical applicability to retrieval tasks on real-world databases is seriously limited by their prohibitively large computational complexity. While dimensionality reduction of the time-series to discrete symbol strings is a standard approach that can exploit computational gains from the data compression as well as the availability of efficient string matching algorithms, the compressed representation of the pitch time series itself is not well understood given the pervasiveness of pitch inflections in the melodic shape of the raga phrases. We propose methods that are informed by domain knowledge to design the representation and to optimize parameter settings for the subsequent string matching algorithm. The methods are evaluated in the context of an audio query based search for Hindustani vocal compositions in audio recordings via the mukhda (refrain of the song). We present results that

demonstrate performance close to that achieved by time-series matching but at orders of magnitude reduction in complexity.

PS3.5 Modified Perceptual Linear Prediction Liftered Cepstrum (MPLPLC) Model for Pop Cover Song Recognition

Ning Chen, J. Stephen Downie, Haidong Xiao, Yu Zhu, Jie Zhu

Most of the features of Cover Song Identification (CSI), for example, Pitch Class Profile (PCP) related features, are based on the musical facets shared among cover versions: melody evolution and harmonic progression. In this work, the perceptual feature was studied for CSI. Our idea was to modify the Perceptual Linear Prediction (PLP) model in the field of Automatic Speech Recognition (ASR) by (a) introducing new research achievements in psychophysics, and (b) considering the difference between speech and music signals to make it consistent with human hearing and more suitable for music signal analysis. Furthermore, the obtained Linear Prediction Coefficients (LPCs) were mapped to LPC cepstrum coefficients, on which liftering was applied, to boost the timbre invariance of the resultant feature: Modified Perceptual Linear Prediction Liftered Cepstrum (MPLPLC). Experimental results showed that both LPC cepstrum coefficients mapping and cepstrum liftering were crucial in ensuring the identification power of the MPLPLC feature. The MPLPLC feature outperformed state-of-the-art features in the context of CSI and in resisting instrumental accompaniment variation. This study verifies that the mature techniques in the ASR or Computational Auditory Scene Analysis (CASA) fields may be modified and included to enhance the performance of the Music Information Retrieval (MIR) scheme.

PS3.6 Raga Verification in Carnatic Music Using Longest Common Segment Set

Shrey Dutta, Krishnaraj Sekhar PV, Hema A. Murthy

There are at least 100 *rāgas* that are regularly performed in Carnatic music concerts. The audience determines the identity of *rāgas* within a few seconds of listening to an item. Most of the audience consists of people who are only avid listeners and not performers.

In this paper, an attempt is made to mimic the listener. A *rāgas* verification framework is therefore suggested. The *rāgas* verification system assumes that a specific *rāgas* is claimed based on similarity of movements and motivic patterns. The system then checks whether this claimed *rāgas* is correct. For every *rāgas*, a set of cohorts are chosen. A *rāgas* and its cohorts are represented using pallavi lines of compositions. A novel approach for matching, called Longest Common Segment Set (LCSS), is introduced. The LCSS scores for a *rāgas* are then normalized with respect to its cohorts in two different ways. The resulting systems and a baseline system are compared for two partitionings of a dataset. A dataset

of 30 *rāgas* from Charsur Foundation 1 is used for analysis. An equal error rate (EER) of 12% is obtained.

PS3.7 Instrument Identification in Optical Music Recognition
Yucong Jiang, Christopher Raphael

We present a method for recognizing and interpreting the text labels for the instruments in an orchestra score, thereby associating staves with instruments. This task is one of many necessary in optical music recognition. Our approach treats the score system as the basic unit of processing. A graph structure describes the possible orderings of instruments in the system. Each instrument may apply to several staves, may be represented with several possible text strings, and may appear at several possible positions relative to the staves. We find the optimal labeling of staves using a globally optimal dynamic programming approach that embeds simple template-based optical character recognition within the overall recognition scheme. When given an entire score, we simultaneously optimize on the text labeling for each system, as well as the character template models, thus adapting to the font at hand. Our implementation alternately optimizes over the text label identification and re-estimates the character templates. Experiments are presented on 10 different scores showing a significant improvement due to adaptation.

PS3.8 Cross-Version Singing Voice Detection in Classical Opera Recordings
Christian Dittmar, Bernhard Lehner, Thomas Prätzlich, Meinard Müller, Gerhard Widmer

In the field of Music Information Retrieval (MIR), the automated detection of the singing voice within a given music recording constitutes a challenging and important research problem. In this study, our goal is to find those segments within a classical opera recording, where one or several singers are active. As our main contributions, we first propose a novel audio feature that extends a state-of-theart feature set that has previously been applied to singing voice detection in popular music recordings. Second, we describe a simple bootstrapping procedure that helps to improve the results in the case that the test data is not reflected well by the training data. Third, we show that a crossversion approach can help to stabilize the results even further.

PS3.9 Accurate Tempo Estimation based on Recurrent Neural Networks and Resonating Comb Filters
Sebastian Böck, Florian Krebs, Gerhard Widmer

In this paper we present a new tempo estimation algorithm which uses a bank of resonating comb filters to determine the dominant periodicity of a musical excerpt. Unlike existing (comb filter based) approaches, we do not use hand-

crafted features derived from the audio signal, but rather let a recurrent neural network learn an intermediate beat-level representation of the signal and use this information as input to the comb filter bank. While most approaches apply complex post-processing to the output of the comb filter bank like tracking multiple time scales, processing different accent bands, modelling metrical relations, categorising the excerpts into slow/ fast or any other advanced processing, we achieve state-of-the-art performance on nine of ten datasets by simply reporting the highest resonator's histogram peak.

PS3.10 Musicology of Early Music with Europeana Tools and Services
Erik Duval, Marnix van Berchum, Anja Jentzsch, Gonzalo Alberto Parra Chico, Andreas Drakos

The Europeana repository hosts large collections of digitized music manuscripts and prints. This paper investigates how tools and services for this repository can enable Early Music musicologists to carry out their research in a more effective or efficient way, or to carry out research that is impossible to do without such tools or services. We report on the methodology, user-centered development of a suite of tools that we have integrated loosely, in order to experiment with this specific target audience and an evaluation of the impact that such tools may have on how these musicologists carry out their research. Positive feedback relates to the automation of data sharing between the loosely coupled tools and support for an integrated workflow. Participants in this study wanted to have the ability to work not only with individual items, but also with collections of such items. The use of search facets to filter, and visualization around time and place were positively evaluated, as was the use of Optical Music Recognition and computer-supported analysis of music scores. The musicologists were not convinced of the value of activity streams. They also wanted a less strictly linear organization of their workflow and the ability to not only consume items from the repository, but to also push their research results back into the Europeana repository.

PS3.11 Singing Voice Separation from Monaural Music Based on Kernel Back-Fitting Using Beta-Order Spectral Amplitude Estimation
Hye-Seung Cho, Jun-Yong Lee, Hyoung-Gook Kim

Separating the leading singing voice from the musical background from a monaural recording is a challenging task that appears naturally in several music processing applications. Recently, kernel additive modeling with generalized spatial Wiener filtering (GW) was presented for music/voice separation. In this paper, an adaptive auditory filtering based on β-order minimum mean-square error spectral amplitude estimation (bSA) is applied to the kernel additive modeling for improving the singing voice separation performance from monaural music signal. The proposed algorithm is composed of five modules: short time Fourier transform,

music/voice separation based on bSA, determination of back-fitting, back-fitting, and inverse short time Fourier transform. In the proposed method, the Singular Value Decomposition (SVD)-based factorized spectral amplitude exponent β for each kernel component is adaptively calculated for effective bSAbased auditory filtering performance during kernel backfitting. Using a back-fitting threshold, the kernel backfitting process can automatically be iteratively performed until convergence. Experimental results show that the proposed method achieves better separation performance than GW based on kernel additive modeling.

PS3.12 Schematizing the Treatment of Dissonance in 16th-Century Counterpoint

Andie Sigler, Jon Wild, Eliot Handelman

We describe a computational project concerning labeling of *dissonance treatments* – schematic descriptions of the uses of dissonances. We use automatic score annotation and database methods to develop schemata for a large corpus of 16th-century polyphonic music. We then apply structural techniques to investigate coincidence of schemata, and to extrapolate from found structures to unused possibilities.

PS3.13 Analysis of the Evolution of Research Groups and Topics in the ISMIR Conference

Mohamed Sordo, Mitsunori Ogihara, Stefan Wuchty

We present an analysis of the topics and research groups that participated in the ISMIR conference over the last 15 years, based on its proceedings. While we first investigate the topological changes of the co-authorship network as well as topics over time, we also identify groups of researchers, allowing us to investigate their evolution and topic dependence. Notably, we find that large groups last longer if they actively alter their membership. Furthermore, such groups tend to cover a wider selection of topics, suggesting that a change of members as well as of research topics increases their adaptability. In turn, smaller groups show the opposite behavior, persisting longer if their membership is altered minimally and focus on a smaller set of topics. Finally, by analyzing the effect of group size and lifespan on research impact, we observed that papers penned by medium sized and long lasting groups tend to have a citation advantage.

PS3.14 A Comparison of Symbolic Similarity Measures for Finding Occurrences of Melodic Segments

Berit Janssen, Peter van Kranenburg, Anja Volk

To find occurrences of melodic segments, such as themes, phrases and motifs, in musical works, a well-performing similarity measure is needed to support human analysis of large music corpora. We evaluate the performance of a range of

melodic similarity measures to find occurrences of phrases in folk song melodies. We compare the similarity measures correlation distance, city-block distance, Euclidean distance and alignment, proposed for melody comparison in computational ethnomusicology; furthermore Implication-Realization structure alignment and B-spline alignment, forming successful approaches in symbolic melodic similarity; moreover, wavelet transform and the geometric approach Structure Induction, having performed well in musical pattern discovery. We evaluate the success of the different similarity measures through observing retrieval success in relation to human annotations. Our results show that local alignment and SIAM perform on an almost equal level to human annotators.

PS3.15 PAD and SAS: Two Awareness-Weighted Rhythmic Similarity Distances

Daniel Gómez-Marín, Sergi Jordà, Perfecto Herrera

Measuring rhythm similarity is relevant for the analysis and generation of music. Existing similarity metrics tend to consider our perception of rhythms as being in time without discriminating the importance of some regions over others. In a previously reported experiment we observed that measures of similarity may differ given the presence or absence of a pulse inducing sound and the importance of those measures is not constant along the pattern. These results are now reinterpreted by refining the previously proposed metrics. We consider that the perceptual contribution of each beat to the measured similarity is non-homogeneous but might indeed depend on the temporal positions of the beat along the bar. We show that with these improvements, the correlation between the previously evaluated experimental similarity and predictions based on our metrics increases substantially. We conclude by discussing a possible new methodology for evaluating rhythmic similarity between audio loops.

PS3.16 Four Timely Insights on Automatic Chord Estimation

Eric J. Humphrey, Juan P. Bello

Automatic chord estimation (ACE) is a hallmark research topic in content-based music informatics, but like many other tasks, system performance appears to be converging to yet another glass ceiling. Looking toward trends in other machine perception domains, one might conclude that complex, data-driven methods have the potential to significantly advance the state of the art. Two recent efforts did exactly this for large-vocabulary ACE, but despite arguably achieving some of the highest results to date, both approaches plateau well short of having solved the problem. Therefore, this work explores the behavior of these two high performing, systems as a means of understanding obstacles and limitations in chord estimation, arriving at four critical observations: one, music recordings that invalidate tacit assumptions about harmony and tonality result in erroneous and

even misleading performance; two, standard lexicons and comparison methods struggle to reflect the natural relationships between chords; three, conventional approaches conflate the competing goals of recognition and transcription to some undefined degree; and four, the perception of chords in real music can be highly subjective, making the very notion of "ground truth" annotations tenuous. Synthesizing these observations, this paper offers possible remedies going forward, and concludes with some perspectives on the future of both ACE research and the field at large.

PS3.17 Improving Melodic Similarity in Indian Art Music Using Culture-Specific Melodic Characteristics

Sankalp Gulati, Joan Serrà, Xavier Serra

Detecting the occurrences of rāgs' characteristic melodic phrases from polyphonic audio recordings is a fundamental task for the analysis and retrieval of Indian art music. We propose an abstraction process and a complexity weighting scheme which improve melodic similarity by exploiting specific melodic characteristics in this music. In addition, we propose a tetrachord normalization to handle transposed phrase occurrences. The melodic abstraction is based on the partial transcription of the steady regions in the melody, followed by a duration truncation step. The proposed complexity weighting accounts for the differences in the melodic complexities of the phrases, a crucial aspect known to distinguish phrases in Carnatic music. For evaluation we use over 5 hours of audio data comprising 625 annotated melodic phrases belonging to 10 different phrase categories. Results show that the proposed melodic abstraction and complexity weighting schemes significantly improve the phrase detection accuracy, and that tetrachord normalization is a successful strategy for dealing with transposed phrase occurrences in Carnatic music. In the future, it would be worthwhile to explore the applicability of the proposed approach to other melody dominant music traditions such as Flamenco, Beijing opera and Turkish Makam music.

PS3.18 Searching Lyrical Phrases in A-Capella Turkish Makam Recordings

Georgi Dzhambazov, Sertan Şentürk, Xavier Serra

Search by lyrics, the problem of locating the exact occurrences of a phrase from lyrics in musical audio, is a recently emerging research topic. Unlike key-phrases in speech, lyrical key-phrases have durations that bear important relation to other musical aspects like the structure of a composition. In this work we propose an approach that address the differences of syllable durations, specific for singing. First a phrase is expanded to MFCC-based phoneme models, trained on speech. Then, we apply dynamic time warping between the phrase and audio to estimate candidate audio segments in the given audio recording. Next, the retrieved audio segments are ranked by means of a novel score-informed hidden Markov

model, in which durations of the syllables within a phrase are explicitly modeled. The proposed approach is evaluated on 12 a-capella audio recordings of Turkish Makam music. Relying on standard speech phonetic models, we arrive at promising results that outperform a baseline approach unaware of lyrics durations. To the best of our knowledge, this is the first work tackling the problem of search by lyrical key-phrases. We expect that it can serve as a baseline for further research on singing material with similar musical characteristics.

PS3.19 Quantifying Lexical Novelty in Song Lyrics
Robert J. Ellis, Zhe Xing, Jiakun Fang, Ye Wang

Novelty is an important psychological construct that affects both perceptual and behavioral processes. Here, we propose a lexical novelty score (LNS) for a song's lyric, based on the statistical properties of a corpus of 275,905 lyrics (available at www.smcnus.org/lyrics/). A lyric-level LNS was derived as a function of the inverse document frequencies of its unique words. An artist-level LNS was then computed using the LNSs of lyrics uniquely associated with each artist. Statistical tests were performed to determine whether lyrics and artists on Billboard Magazine's lists of "All-Time Top 100" songs and artists had significantly lower LNSs than "non-top" songs and artists. An affirmative and highly consistent answer was found in both cases. These results highlight the potential utility of the LNS as a feature for MIR.

PS3.20 An Efficient Temporally-Constrained Probabilistic Model for Multiple-Instrument Music Transcription
Emmanouil Benetos, Tillman Weyde

In this paper, an efficient, general-purpose model for multiple instrument polyphonic music transcription is proposed. The model is based on probabilistic latent component analysis and supports the use of sound state spectral templates, which represent the temporal evolution of each note (e.g. attack, sustain, decay). As input, a variable-Q transform (VQT) time-frequency representation is used. Computational efficiency is achieved by supporting the use of preextracted and pre-shifted sound state templates. Two variants are presented: without temporal constraints and with hiddenMarkovmodel-based constraints controlling the appearance of sound states. Experiments are performed on benchmark transcription datasets: MAPS, TRIOS,MIREX multiF0, and Bach10; results on multi-pitch detection and instrument assignment show that the proposed models outperform the state-of-the-art for multiple-instrument transcription and is more than 20 times faster compared to a previous sound state-based model. We finally show that

a VQT representation can lead to improved multi-pitch detection performance compared with constant-Q representations.

PS3.21 Electric Guitar Playing Technique Detection in Real-World Recordings Based on F0 Sequence Pattern Recognition
Yuan-Ping Chen, Li Su, Yi-Hsuan Yang

For a complete transcription of a guitar performance, the detection of playing techniques such as bend and vibrato is important, because playing techniques suggest how the melody is interpreted through the manipulation of the guitar strings. While existing work mostly focused on playing technique detection for individual single notes, this paper attempts to expand this endeavor to recordings of guitar solo tracks. Specifically, we treat the task as a time sequence pattern recognition problem, and develop a twostage framework for detecting five fundamental playing techniques used by the electric guitar. Given an audio track, the first stage identifies prominent candidates by analyzing the extracted melody contour, and the second stage applies a pre-trained classifier to the candidates for playing technique detection using a set of timbre and pitch features. The effectiveness of the proposed framework is validated on a new dataset comprising of 42 electric guitar solo tracks without accompaniment, each of which covers 10 to 25 notes. The best average F-score achieves 74% in two-fold cross validation. Furthermore, we also evaluate the performance of the proposed framework for bend detection in five studio mixtures, to discuss how it can be applied in transcribing real-world electric guitar solos with accompaniment.

PS3.22 Extending a Model of Monophonic Hierarchical Music Analysis to Homophony
Phillip B. Kirlin, David L. Thomas

Computers are now powerful enough and data sets large enough to enable completely data-driven studies of Schenkerian analysis, the most well-established variety of hierarchical music analysis. In particular, we now have probabilistic models that can be trained via machine learning algorithms to analyze music in a hierarchical fashion as a music theorist would. Most of these models, however, only analyze the monophonic melodic content of the music, as opposed to taking all of the musical voices into account. In this paper, we explore the feasibility of extending a probabilistic model developed for analyzing monophonic music to function with homophonic music. We present details of the new model, an algorithm for determining the most probable analysis of the music, and a number of experiments evaluating the quality of the analyses predicted by the model. We also describe how varying the way the model interprets rests in the input music affects the resulting analyses produced.

PS3.23 The MIR Perspective on the Evolution of Dynamics in Mainstream Music

Emmanuel Deruty, François Pachet

Understanding the evolution of mainstream music is of high interest for the music production industry. In this context, we argue that a MIR perspective may be used to highlight, in particular, relations between dynamics and various properties of mainstream music. We illustrate this claim with two results obtained from a diachronic analysis performed on 7200 tracks released between 1967 and 2014. This analysis suggests that 1) the so-called "loudness war" has peaked in 2007, and 2) its influence has been important enough to override the impact of genre on dynamics. In other words, dynamics in mainstream music are primarily related to a track's year of release, rather than to its genre.

PS3.24 Theme and Variation Encodings with Roman Numerals (TAVERN): A New Data Set for Symbolic Music Analysis

Johanna Devaney, Claire Arthur, Nathaniel Condit-Schultz, Kirsten Nisula

The Theme And Variation Encodings with Roman Numerals (TAVERN) dataset consists of 27 complete sets of theme and variations for piano composed between 1765 and 1810 by Mozart and Beethoven. In these theme and variation sets, comparable harmonic structures are realized in different ways. This facilitates an evaluation of the effectiveness of automatic analysis algorithms in generalizing across different musical textures. The pieces are encoded in standard **kern format, with analyses jointly encoded using an extension to **kern. The harmonic content of the music was analyzed with both Roman numerals and function labels in duplicate by two different expert analyzers. The pieces are divided into musical phrases, allowing for multiple-levels of automatic analysis, including chord labeling and phrase parsing. This paper describes the content of the dataset in detail, including the types of chords represented, and discusses the ways in which the analyzers sometimes disagreed on the lower-level harmonic content (the Roman numerals) while converging at similar high-level structures (the function of the chords within the phrase).

PS3.25 Benford's Lawfor Music Analysis

Isabel Barbancho, Lorenzo J. Tardón, Ana M. Barbancho, Mateu Sbert

Benford's law defines a peculiar distribution of the leading digits of a set of numbers. The behavior is logarithmic, with the leading digit 1 reflecting largest probability of occurrence and the remaining ones showing decreasing probabilities of appearance following a logarithmic trend. Many discussions have been carried out about the application of Benford's law tomany different fields. In this paper, a novel exploitation of Benford's law for the analysis of audio signals is

proposed. Three new audio features based on the evaluation of the degree of agreement of a certain audio dataset to Benford's law are presented. These new proposed features are succesfully tested in two concrete audio tasks: the detection of artificially assembled chords and the estimation of the quality of the MIDI conversions.

PS3.26 An Audio to Score Alignment Framework Using Spectral Factorization and Dynamic Timewarping

J.J. Carabias-Orti, F.J. Rodriguez-Serrano, P. Vera-Candeas, N. Ruiz-Reyes, F.J. Cañadas-Quesada

In this paper, we present an audio to score alignment framework based on spectral factorization and online Dynamic Time Warping (DTW). The proposed framework has two separated stages: preprocessing and alignment. In the first stage, we use Non-negative Matrix Factorization (NMF) to learn spectral patterns (i.e. basis functions) associated to each combination of concurrent notes in the score. In the second stage, a low latency signal decomposition method with fixed spectral patterns per combination of notes is used over the magnitude spectrogram of the input signal resulting in a divergence matrix that can be interpreted as the cost of the matching for each combination of notes at each frame. Finally, a Dynamic Time Warping (DTW) approach has been used to find the path with the minimum cost and then determine the relation between the performance and the musical score times. Our framework have been evaluated using a dataset of baroque-era pieces and compared to other systems, yielding solid results and performance.

PS3.27 New Sonorities for Early Jazz Recordings Using Sound Source Separation and Automatic Mixing Tools

Daniel Matz, Estefanía Cano, Jakob Abeßer

In this paper, a framework for automatic mixing of early jazz recordings is presented. In particular, we propose the use of sound source separation techniques as a preprocessing step of the mixing process. In addition to an initial solo and accompaniment separation step, the proposed mixing framework is composed of six processing blocks: harmonic-percussive separation (HPS), cross-adaptive multitrack scaling (CAMTS), cross-adaptive equalizer (CAEQ), cross-adaptive dynamic spectral panning (CADSP), automatic excitation (AE), and time-frequency selective panning (TFSP). The effects of the different processing steps in the final quality of the mix are evaluated through a listening test procedure. The results show that the desired quality improvements in terms of sound balance, transparency, stereo impression, timbre, and overall impression can be achieved with the proposed framework.

PS3.28 Automatic Transcription of Ornamented Irish Traditional Flute Music Using Hidden Markov Models

Peter Jančovič, Münevver Köküer, Wrena Baptiste

This paper presents an automatic system for note transcription of Irish traditional flute music containing ornamentation. This is a challenging problem due to the soft nature of onsets and short durations of ornaments. The proposed automatic transcription system is based on hidden Markov models, with separate models being built for notes and for single-note ornaments. Mel-frequency cepstral coefficients are employed to represent the acoustic signal. Different setups of parameters in feature extraction and acoustic modelling are explored. Experimental evaluations are performed on monophonic flute recordings from Grey Larsen's CD. The performance of the system is evaluated in terms of the transcription of notes as well as detection of onsets. It is demonstrated that the proposed system can achieve a very good note transcription and onset detection performance. Over 28% relative improvement in terms of the F-measure is achieved for onset detection in comparison to conventional onset detection methods based on signal energy and fundamental frequency.

PS3.29 Towards Music Imagery Information Retrieval: Introducing the OpenMIIR Dataset of EEG Recordings from Music Perception and Imagination

Sebastian Stober, Avital Sternin, Adrian M. Owen, Jessica A. Grahn

Music imagery information retrieval (MIIR) systems may one day be able to recognize a song from only our thoughts. As a step towards such technology, we are presenting a public domain dataset of electroencephalography (EEG) recordings taken during music perception and imagination. We acquired this data during an ongoing study that so far comprises 10 subjects listening to and imagining 12 short music fragments – each 7–16s long – taken from well-known pieces. These stimuli were selected from different genres and systematically vary along musical dimensions such as meter, tempo and the presence of lyrics. This way, various retrieval scenarios can be addressed and the success of classifying based on specific dimensions can be tested. The dataset is aimed to enable music information retrieval researchers interested in these newMIIR challenges to easily test and adapt their existing approaches for music analysis like fingerprinting, beat tracking, or tempo estimation on EEG data.

PS3.30 Emotion Based Segmentation of Musical Audio

Anna Aljanaki, Frans Wiering, Remco C. Veltkamp

The dominant approach to musical emotion variation detection tracks emotion over time continuously and usually deals with time resolutions of one second. In this paper we discuss the problems associated with this approach and propose to move to bigger time resolutions when tracking emotion over time. We argue that it is more natural from the listener's point of view to regard emotional variation in music as a progression of emotionally stable segments. In order to enable such tracking of emotion over time it is necessary to segment music at the emotional boundaries. To address this problem we conduct a formal evaluation of different segmentation methods as applied to a task of emotional boundary detection. We collect emotional boundary annotations from three annotators for 52 musical pieces from the RWC music collection that already have structural annotations from the SALAMI dataset. We investigate how well structural segmentation explains emotional segmentation and find that there is a large overlap, though about a quarter of emotional boundaries do not coincide with structural ones. We also study inter-annotator agreement on emotional segmentation. Lastly, we evaluate different unsupervised segmentation methods when applied to emotional boundary detection and find that, in terms of F-measure, the Structural Features method performs best.

Oral Session 6
User & Communnity

14:30-16:00 Thursday, October 29

Session Chair: Yi-Hsuan Yang

OS6.1 MIREX Grand Challenge 2014 User Experience: Qualitative Analysis of User Feedback

Jin Ha Lee, Xiao Hu, Kahyun Choi, J. Stephen Downie

Evaluation has always been fundamental to the Music Information Retrieval (MIR) community, as evidenced by the popularity of the Music Information Retrieval Evaluation eXchange (MIREX). However, prior MIREX tasks have primarily focused on testing specialized MIR algorithms that sit on the back end of systems. Not until the Grand Challenge 2014 User Experience (GC14UX) task had the users' overall interaction and experience with complete systems been formally evaluated. Three systems were evaluated based on five criteria. This paper reports the results of GC14UX, with a special focus on the qualitative analysis of 99 free text responses collected from evaluators. The analysis revealed additional user opinions, not fully captured by score ratings on the given criteria, and demonstrated the challenge of evaluating a variety of systems with different user goals. We conclude with a discussion on the implications of findings and recommendations for future UX evaluation tasks, including adding new criteria: Aesthetics, Performance, and Utility.

OS6.2 AcousticBrainz: A Community Platform for Gathering Music Information Obtained from Audio

Alastair Porter, Dmitry Bogdanov, Robert Kaye, Roman Tsukanov, Xavier Serra

We introduce the AcousticBrainz project, an open platform for gathering music information. At its core, AcousticBrainz is a database of music descriptors computed from audio recordings using a number of state-of-the-art Music Information Retrieval algorithms. Users run a supplied feature extractor on audio files and upload the analysis results to the AcousticBrainz server. All submissions include a MusicBrainz identifier allowing them to be linked to various sources of editorial information. The feature extractor is based on the open source Essentia audio analysis library. From the data submitted by the community, we run classifiers aimed at adding musically relevant semantic information. These classifiers can be developed by the community using tools available on the AcousticBrainz website. All data in AcousticBrainz is freely available and can be accessed through the website or API. For AcousticBrainz to be successful we need to have an active community that contributes to and uses this platform, and it is this community that will define the actual uses and applications of its data.

OS6.3 How Music Alters Decision Making - Impact of Music Stimuli on Emotional Classification

Elad Liebman, Peter Stone, Corey N. White

Numerous studies have demonstrated that mood can affect emotional processing. The goal of this study was to explore which components of the decision process are affected when exposed to music; we do so within the context of a stochastic sequential model of simple decisions, the drift-diffusion model (DDM). In our experiment, participants decided whether words were emotionally positive or negative while listening to music that was chosen to induce positive or negative mood. The behavioral results show that the music manipulation was effective, as participants were biased to label words positive in the positive music condition. The DDM shows that this bias was driven by a change in the starting point of evidence accumulation, which indicates an a priori response bias. In contrast, there was no evidence that music affected how participants evaluated the emotional content of the stimuli. To better understand the correspondence between auditory features and decision-making, we proceeded to study how individual aspects of music affect response patterns. Our results have implications for future studies of the connection between music and mood.

OS6.4 Put the Concert Attendee in the Spotlight. A User-Centered Design and Development Approach for Classical Concert Applications

Mark S. Melenhorst, Cynthia C. S. Liem

As the importance of real-life use cases in the music information retrieval (MIR) field is increasing, so does the importance of understanding user needs. The development of innovative real-life applications that draw on MIR technology requires a user-centered design and development approach that assesses user needs and aligns them with technological and academic ambitions in the MIR domain. In this paper we present such an approach, and apply it to the development of technological applications to enrich classical symphonic concerts. A userdriven approach is particularly important in this area, as orchestras need to innovate the concert experience to meet the needs and expectations of younger generations without alienating the current audience. We illustrate this approach with the results of five focus groups for three audience segments, which allow us to formulate informed user requirements for classical concert applications.

Oral Session 7
Performance

9:00-10:00 Friday, October 30

Session Chair: Zhiyao Duan

OS7.1 Analysis of Expressive Musical Terms in Violin Using Score-Informed and Expression-Based Audio Features

Pei-Ching Li, Li Su, Yi-Hsuan Yang, Alvin W. Y. Su

The manipulation of different interpretational factors, including dynamics, duration, and vibrato, constitutes the realization of different expressions in music. Therefore, a deeper understanding of the workings of these factors is critical for advanced expressive synthesis and computeraided music education. In this paper, we propose the novel task of automatic expressive musical term classification as a direct means to study the interpretational factors. Specifically, we consider up to 10 expressive musical terms, such as Scherzando and Tranquillo, and compile a new dataset of solo violin excerpts featuring the realization of different expressive terms by different musicians for the same set of classical music pieces. Under a score-informed scheme, we design and evaluate a number of note-level features characterizing the interpretational aspects of music for the classification task. Our evaluation shows that the proposed features lead to significantly higher classification accuracy than a baseline feature set commonly used in music information retrieval tasks. Moreover, taking the contrast of feature values between an expressive and its corresponding non-expressive version (if given) of a music piece greatly improves the accuracy in classifying the presented expressive one. We also draw insights from analyzing the feature relevance and the class-wise accuracy of the prediction.

OS7.2 Spectral Learning for Expressive Interactive Ensemble Music Performance

Guangyu Xia, Yun Wang, Roger B. Dannenberg, Geoffrey Gordon

We apply machine learning to a database of recorded ensemble performances to build an artificial performer that can perform music expressively in concert with human musicians. We consider the piano duet scenario and focus on the interaction of expressive timing and dynamics. We model different performers' musical expression as coevolving time series and learn their interactive relationship from multiple rehearsals. In particular, we use a spectral method, which is able to learn the correspondence not only between different performers but also between the performance past and future by reduced-rank partial regressions. We describe our model that captures the intrinsic interactive relationship between different performers, present the spectral learning procedure, and show that the spectral learning algorithm is able to generate a more human-like interaction.

OS7.3 Score-Informed Analysis of Intonation and Pitch Modulation in Jazz Solos

Jakob Abeßer, Estefanía Cano, Klaus Frieler, Martin Pfleiderer, Wolf-Georg Zaddach

The paper presents new approaches for analyzing the characteristics of intonation and pitch modulation of woodwind and brass solos in jazz recordings. To this end, we use score-informed analysis techniques for source separation and fundamental frequency tracking. After splitting the audio into a solo and a backing track, a reference tuning frequency is estimated from the backing track. Next, we compute the fundamental frequency contour for each tone in the solo and a set of features describing its temporal shape. Based on this data, we first investigate, whether the tuning frequencies of jazz recordings changed over the decades of the last century. Second, we analyze whether the intonation is artist-specific. Finally, we examine how the modulation frequency of vibrato tones depends on contextual parameters such as pitch, duration, and tempo as well as the performing artist.

Acknowledgments

The 16th International Society for Music Information Retrieval Conference (ISMIR 2015) was made possible thanks to the hard work of many people including the many reviewers, meta-reviewers and all the members of the Conference Committee. Special thanks go to this year's sponsors:

- Gold Sponsors:
  - Fundación Española para la Ciencia y la Tecnología (FECYT), Ministerio de Economía y Competitividad, Gobierno de España (under project FCT-14-8217)
  - Shazam
  - Gracenote
  - Pandora
  - Bose Corporation
  - Universidad de Málaga, Vicerrectorado de Investigación y Transferencia, Campus de Excelencia Internacional Andalucia Tech.
  - Ministerio de Economía y Competitividad, Gobierno de España (under project TIN2015-62946-CIN)
- Silver Sponsors:
  - Steinberg Media Technologies GmbH
  - Native Instruments GmbH
- Bronze Sponsors:
  - Google, Inc
  - ACRCloud - Automatic Content Recognition Cloud Service
  - Smule

- Other Collaborators:
  - Departamento de Ingeniería de Comunicaciones, Universidad de Málaga
  - E.T.S. Ingeniería de Telecomunicación, Universidad de Málaga
  - Fundación Unicaja
  - Ayuntamiento de Málaga
  - Málaga Convention Bureau

Last, but not least, the ISMIR program is possible only thanks to the excellent contributions of the ISMIR community in response to our call for participation. The biggest acknowledgment goes to you, the authors, researchers and participants of this conference. We hope you all have a wonderful and unforgettable stay in Spain!

Isabel Barbancho
Lorenzo J. Tardón
General Chairs, ISMIR 2015

Meinard Müller
Frans Wiering
Program Chairs, ISMIR 2015

Reviewers

Abesser, Jakob
Ahonen, Teppo
Albrecht, Joshua
Aljanaki, Anna
Anden, Joakim
Arjannikov, Tom
Arzt, Andreas

Bailes, Freya
Bainbridge, David
Balke, Stefan
Barbancho, Ana María
Barbancho, Isabel
Basaran, Dogaç
Baumann, Stephan
Bay, Mert
Bazzica, Alessio
Bellogín, Alejandro
Bemman, Brian
Bennett, Christopher
Bernardes, Gilberto
Bigo, Louis
Bimbot, Frédéric
Bittner, Rachel
Bohak, Ciril
Bonafonte, Antonio
Bosch, Juanjo
Bountouridis, Dimitrios

Caetano, Marcelo
Cambouropoulos, Emilios
Cancino Chacon, Carlos E.
Cano, Estefanía
Carabias, Julio
Cartwright, Mark
Chan, Tak-Shing
Chen, Shuo
Chen, Chih-Ming
Cherla, Srikanth
Chin, Yu-Hao
Choi, Kahyun
Chung, Chia-Hao
Cogliati, Andrea
Collins, Nick
Cont, Arshia
Cotton, Courtenay
Coviello, Emanuele
Crawford, Tim

David, Bertrand
De Man, Brecht
Degara, Norberto
Dessein, Arnaud
Di Giorgi, Bruno
Driedger, Jonathan
Durand, Simon

Eghbal-Zadeh, Hamid
Elowsson, Anders

Essinger, Steve

Faraldo, Ángel
Fernandes Tavares, Tiago
Fields, Ben
Fillon, Thomas
Fitzgerald, Derry
Fonseca, Nuno
Font, Frederic
Foster, Peter
Fukayama, Satoru

Ganseman, Joachim
Gedik, Ali Cenk
Gingras, Bruno
Giraud, Mathieu
Goebl, Werner
Gómez, Daniel
Grill, Thomas
Grohganz, Harald
Grunberg, David
Guastavino, Catherine
Gulati, Sankalp
Gurevich, Michael

Hahn, Henrik
Hamanaka, Masatoshi
Hamel, Philippe
Hankinson, Andrew
Hanna, Pierre
Hargreaves, Steven
Hedges, Tom
Herrera, Jorge
Hoover, Amy
Huang, Anna
Huang, Po-Sen

Imbrasaite, Vaiva
Ishwar, Vignesh
Itoyama, Katsutoshi

Jakubowski, Kelly
Janssen, Berit
Jao, Ping-Keng
Jensen, Kristoffer
Jordà, Sergi

Kaneshiro, Blair
Keller, Damián
Kelz, Rainer
Kim, Minje
Klapuri, Anssi
Knopke, Ian
Koerich, Alessandro
Koops, Vincent
Korzeniowski, Filip
Kosta, Katerina
Krebs, Florian
Kroher, Nadine
Kruspe, Anna
Ku, Lun-Wei

Lagrange, Mathieu
Langlois, Thibault
Lartillot, Olivier
Lattner, Stefan
Lee, Kyogu
Lemstrom, Kjell
Leve, Florence
Levy, Mark
Lewis, David
Lewis, Richard
Li, Bochen
Li, Wei
Liang, Dawen
Lidy, Thomas
Lin, Yin-Tzu
Liu, Yi-Wen
Liu, Jen-Yu
Liutkus, Antoine
Loviscach, Jörn
Lu, Lie
Lukashevich, Hanna

Machado, Anderson

Magalhẽs, José Pedro
Mandel, Michael
Margulis, Elizabeth
Marolt, Matija
Marrero, Mónica
Marsden, Alan
Martorell, Agustin
Mayer, Rudolf
McFee, Brian
McKay, Cory
Meseguer, Gabriel
Mesnage, Cedric
Milne, Andrew
Miron, Marius
Moore, Josh
Morvidone, Marcela
Moussallam, Manuel
Mü llensiefen, Daniel

Nagano, Hidehisa
Nakano, Tomoyasu
Nakano, Masahiro
Nam, Juhan
Naveda, Luiz
Neubarth, Kerstin
Nichols, Eric
Nieto, Oriol

Ogihara, Mitsunori
Oland, Anders
Oramas, Sergio
Ostuni, Vito

Page, Kevin
Panteli, Maria
Papadopoulos, Hélène
Paulus, Jouni
Pauwels, Johan
Percival, Graham
Perez, Alfonso
Pertusa, Antonio
Pesek, Matevž
Pikrakis, Aggelos
Prätzlich, Thomas
Preinfalk, Martin
Prockup, Matthew
Proutskova, Polina
Pugin, Laurent

Quinton, Elio

Raczynski, Stanislaw
Raffel, Colin
Rafii, Zafar
Ramírez, Miguel
Rao, Preeti
Rashkov, Valery
Rauber, Andreas
Rebelo, Ana
Ren, Gang
Ringwalt, Daniel
Rizo, David
Robertson, Andrew
Robine, Matthias
Rocamora, Martin
Rocha, Bruno
Rodríguez Algarra, Francisco
Rodríguez López, Marcelo
Roma, Gerard
Rowe, Robert
Russo, Frank

Saari, Pasi
Sadakata, Makiko
Said, Alan
Schindler, Alexander
Schlüter, Jan
Schuller, Björn
Seetharaman, Prem
Sentürk, Sertan
Sigler, Andie
Sigtia, Siddharth Sanjay
Smaragdis, Paris
Smith, Lloyd

Soleymani, Mohammad
Song, Yading
Sonnleitner, Reinhard
Souza Britto Jr., Alceu de
Sprechmann, Pablo
Srinivasamurthy, Ajay
Stewart, Rebecca
Stöter, Fabian-Robert
Stowell, Dan
Subramanian, Anand
Summers, Cameron
Supper, Ben

Temperley, David
Thalmann, Florian
Tian, Mi
Tjoa, Steve
Tkalcic, Marko
Toussaint, Godfried

Uhle, Christian
Upham, Finn

Valk, Reinier de
Vall, Andreu
Valle, Rafael
Van Balen, Jan
Van Handel, Leigh
Velarde, Gissel
Veltkamp, Remco
Vempala, Naresh
Vigliensoni, Gabriel
Vogl, Richard

Wang, Xing
Wang, Siying
Wang, Xinxi
Wang, Hsin-Min
Weerkamp, Wouter
Weigl, David
Weij, Bastiaan van der
Weiss, Ron
Weiß, Christof
Whalley, Ian
Wolff, Daniel
Wu, Fu-Hai Frank
Wu, Ben

Xia, Guangyu

Zacharakis, Asteris
Zanoni, Massimiliano
Zbikowski, Lawrence

Author Index

Notes

ORGANIZATION

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

UNIVERSIDAD DE MÁLAGA

Escuela Técnica Superior de Ingeniería de Telecomunicación

ISMIR

Gold Sponsors

GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

FECYT FUNDACIÓN ESPAÑOLA PARA LA CIENCIA Y LA TECNOLOGÍA

shazam

gracenote

PANDORA

BOSE

UNIVERSIDAD DE MÁLAGA

ANDALUCÍA TECH

Vicerrectorado de Investigación y Transferencia

GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Silver Sponsors

steinberg

NATIVE INSTRUMENTS

Bronze Sponsors

Google

ACRCloud

smule

Other Collaborators

Ayuntamiento de Málaga

Málaga Convention Bureau

Unicaja Fundación

ISMIR 2015 MALAGA

16th International Society for Music Information Retrieval Conference
26 - 30 October 2015, Málaga (Spain)

Conference Program

<table>
<tr><th>WHEN & WHERE</th><th>MONDAY, OCTOBER 26 E.T.S.I.T</th><th>TUESDAY, OCTOBER 27 NH</th><th>WEDNESDAY, OCTOBER 28 NH</th><th>THURSDAY, OCTOBER 29 NH</th><th>FRIDAY, OCTOBER 30 NH</th></tr>
<tr><td>9:00</td><td>Registration</td><td>Registration</td><td>Oral Session 2</td><td>Oral Session 5</td><td>Oral Session 7</td></tr>
<tr><td>10:00</td><td rowspan="3">Tutorials 1, 2 and 3</td><td>Opening</td><td rowspan="4">Poster Session 2 & Coffee</td><td rowspan="4">Poster Session 3 & Coffee</td><td rowspan="2">MIREX Oral Session</td></tr>
<tr><td>10:30</td><td rowspan="2">Keynote Speaker 1</td></tr>
<tr><td>11:00</td><td rowspan="3">MIREX Posters & Coffee</td></tr>
<tr><td>11:30</td><td>Coffee</td><td rowspan="3">Poster Session 1 & Coffee</td></tr>
<tr><td>12:00</td><td rowspan="2">Tutorials 1, 2 and 3</td><td>Oral Session 3</td><td>Keynote Speaker 2</td></tr>
<tr><td>12:30</td><td>WiMIR</td><td>Industrial Panel</td><td>Business Meeting</td></tr>
<tr><td>13:30</td><td rowspan="2">Break</td><td rowspan="2">Lunch</td><td rowspan="2">Lunch</td><td rowspan="2">Lunch</td><td>Closing</td></tr>
<tr><td>13:40</td><td>Lunch</td></tr>
<tr><td>14:30</td><td>Tutorials 4, 5 and 6</td><td>Oral Session 1</td><td>Oral Session 4</td><td>Oral Session 6</td><td>Demos & Late-Breaking News</td></tr>
<tr><td>16:00</td><td>Coffee</td><td rowspan="2">Poster Session 1 & Coffee</td><td rowspan="2">Poster Session 2 & Coffee</td><td rowspan="2">Poster Session 3 & Coffee</td><td rowspan="2">Unconference</td></tr>
<tr><td>16:30</td><td rowspan="2">Tutorials 4, 5 and 6</td></tr>
<tr><td>17:30</td><td></td><td></td><td></td><td></td></tr>
<tr><td>18:00</td><td></td><td></td><td rowspan="3">Flamenco Show & Concert at Conservatorio María Cristina</td><td></td><td></td></tr>
<tr><td>18:30</td><td></td><td></td><td></td><td></td></tr>
<tr><td>20:00</td><td rowspan="2">Welcome Reception</td><td></td><td rowspan="2">Gala Dinner & ISMIR 2015 Pandora Jam Session</td><td></td></tr>
<tr><td>22:00</td><td></td><td></td><td></td></tr>
</table>