applied
sciences

MDPI

Meeting Report

16th Sound and Music Computing Conference SMC 2019 (28–31 May 2019, Malaga, Spain)

Lorenzo J. Tardón [1,*], **Isabel Barbancho** [1,*], **Ana M. Barbancho** [1], **Alberto Peinado** [2], **Stefania Serafin** [3] **and Federico Avanzini** [4]

1   ATIC Research Group, Universidad de Málaga, Andalucía Tech, E.T.S.I. Telecomunicación, 29071 Málaga, Spain; abp@ic.uma.es
2   Universidad de Málaga, Andalucía Tech, E.T.S.I. Telecomunicación, 29071 Málaga, Spain; apd@uma.es
3   Multisensory Experience Lab., Aalborg University Copenhagen, 2450 Copenhagen SV, Denmark; sts@create.aau.dk
4   LIM–Laboratorio di Informatica Musicale, Department of Computer Science, University of Milan, 20133 Milan, Italy; federico.avanzini@di.unimi.it
*   Correspondence: lorenzo@ic.uma.es (L.J.T.); ibp@ic.uma.es (I.B.)

check for updates

Abstract: The **16th Sound and Music Computing Conference** (**SMC 2019**) took place in Malaga, Spain, 28–31 May 2019 and it was organized by the Application of Information and Communication Technologies Research group (ATIC) of the University of Malaga (UMA). The **SMC 2019 associated Summer School** took place 25–28 May 2019. The **First International Day of Women in Inclusive Engineering, Sound and Music Computing Research** (**WiSMC 2019**) took place on 28 May 2019. The **SMC 2019 TOPICS OF INTEREST** included a wide selection of topics related to acoustics, psychoacoustics, music, technology for music, audio analysis, musicology, sonification, music games, machine learning, serious games, immersive audio, sound synthesis, etc.

1. Summer School

1.1. Arduino and Audio

David Cuartielles
Malmö University, Sweden

The Arduino and audio workshop looks at possible ways to create inter- active sound production machines using Arduino boards. From low level bit- banging PCM on Arduino UNO as a way to make inexpensive sound toys, to the use of DAC in more modern processors. The workshop has a strong hands-on component where all participants will make a small music instrument (or two) using 8-bit processors.

1.2. Optical Music Recognition (OMR)

Jorge Calvo-Zaragoza
University of Alicante, Spain

Optical Music Recognition (OMR) is the field that investigates how to teach computers to read musical scores. Bringing OMR to real use lowers the costs of making written music available in symbolic format, diversifying the sources for music information retrieval and digital musicology. State-of-the-art research and existing tools are described so that attendees can integrate OMR into their own work.

1.3. Wiring, Soldering and Enclosing Music

Koka Nikoladze
Norwegian Academy of Music, Norway

This one-day interactive workshop focuses on different aspects of engineering and crafting expressive musical interfaces for specific musical works. Koka presents and discusses some of his inventions and designs. He also provides an in-depth view in his composition and engineering toolkits and workflows. Controlling a symphonic orchestra in realtime, performing with a YouTube choir, playing a laptop with a viola bow—the workshop is planned to be highly informative, but also entertaining.

1.4. Music Recommendation

Peter Knees
Faculty of Informatics, TU Wien, Austria

During this one-day course, the students learn about different types of recommenders, user aspects, recommenders for creators, and even some business aspects of it. The user aspects mainly focus on interaction.

2. First International Day of Women in Inclusive Engineering, Sound and Music Computing Research WiSMC 2019

This day is specially addressed to pre-university students between 15 and 17 years old that have to decide what they want to study after finishing high school. To make them aware that they can study anything they want and engineering is a good option no matter if you are a woman or a man.

Therefore, the aim of WiSMC 2019 is to realize that it is necessary that both women and men work in collaboration, supporting each other, so that we all can reach fulfillment as individuals. With this goal in mind, women with successful professional careers who have also formed a family are invited to tell their life experience. Thus, girls, young women, and all the assistants can realize that women do not have to give up anything and that, with the adequate support from the rest of society, it is possible to achieve any kind of goal.

WiSMC 2019 Invited Speakers

Elvira Brattico

Principal Investigator at the Center for Music in the Brain, a center of excellence funded by the Danish National Research Foundation and affiliated with the Department of Clinical Medicine at Aarhus University and The Royal Academy of Music Aarhus/Aalborg, Aarhus, Denmark.

Stefania Serafin

Professor at Aalborg University in Copenhagen. The president of the Sound and Music Computing Association and Project Leader for the Nordic Sound and Music Computing Network.

Anja Volk

Assistant Professor in Information and Computing Sciences at Utrecht University, the Netherlands; with a dual background in mathematics and musicology, she has an international reputation in the areas of music information retrieval (MIR), computational musicology, and mathematical music theory.

Ana Rivera

Telecommunication Engineer from University of Málaga. She is with the Program Office for the Development of 5G Technology project at the WDO division of Keysight Technologies.

Gema Martín

She holds a Mathematics Degree, a Software Engineer PhD. and Master Degrees in different areas: Technological and Innovative Entrepreneurs Master, Software Engineering and Artificial Intelligence Master, Social Networks Master and Higher Educational Teaching Master. Currently, she is R&D Manager at AGANOVA.

Ana Pedraz

Snr Manager Solutions Engineer Iberia–Oracle. She also leads the Oracle Woman Leadership community and coaches their interns thru an innovation program.

3. Keynote Talks

3.1. Taming the Untameable: How to Study Naturalistic Music Listening in the Brain by Means of Computational Feature Extraction

Elvira Brattico
Dept. Clinical Medicine—Center for Music. In the Brain, Aarhus C, Denmark

Listening to musical sounds is a brain function that has likely appeared already tens of thousands of years ago, in *Homo sapiens* and perhaps even in Neanderthal ancestors. The peripheral hearing apparatus has taken its shape to decompose sounds by transforming air pressure waves into ion impulses and by extracting the frequencies in a way similar to Fourier transform at the level of the basilar membrane in the inner ear. These neuronal codes are then transferred in several relay stations of the central nervous system to reach the primary and non-primary auditory cerebral cortex. The ways those codes for musical sounds are obtained and represented in the cerebral cortex is only partially understood. To investigate this, various stimulation paradigms have been developed, most of them being distant from naturalistic constantly-varying sound environments in order to maintain strict control over manipulated variables. This controlled approach limits the generalization of findings to real-life listening situations. In our recent studies, we introduced a novel experimental paradigm where participants are simply asked to naturalistically listen to music rather than to perform tasks in response to some artificial sounds. This free-listening paradigm benefits from music information retrieval, since it handles the computationally extracted features from the music as time series variables to be related to the brain signal. Our studies have advanced the understanding of music processing in the brain, demonstrating activity in large-scale networks connecting audio-motor, emotion, and cognitive regions of the brain during the act of listening to whole pieces of music.

3.2. Towards Explicating Implicit Musical Knowledge: How the Computational Modeling of Musical Structures Mediates between Curiosity-Driven and Application—Oriented Perspectives

Anja Volk
Dept. Information and Computing Sciences, Utrecht University, Netherlands

Over the past decades, we have witnessed a rapid development of music technology for many different application contexts, such as music recommender systems, music search engines, automatic music generation systems, and new interactive musical instruments. They have enabled new ways of accessing and interacting with music. At the same time, the process of developing these new technologies employing an application-oriented perspective has revealed many open questions about music as a fundamental human trait. In this talk, I will discuss how the explicit modeling of musical structures in the computational domain uncovers layers of implicit musical knowledge applied by expert and ordinary listeners when interacting with music. Starting from our research on developing online search methods for Dutch folk songs and on developing online music education systems, I will demonstrate how crucial concepts, such as music similarity, harmonic variance, and repeated patterns, are scrutinized in the process of developing computational models. The explicit modeling within the

computational context enhances our understanding of how we employ these concepts implicitly when interacting with music. This contributes to curiosity-driven research about music as a fundamental human trait, paving the way for cross-disciplinary approaches to music encompassing computer science, musicology, and cognition.

3.3. Music's Changing Fast; FAST Is Changing Music

Mark Sandler
Centre for Digital Music, Queen Mary University of London, UK

The FAST project (Fusing Audio and Semantic Technology for Intelligent Music Production and Consumption), with 5 years of UK funding, seeks to create a new technological ecosystem for recorded music that empowers people throughout the value chain, from professional performers to casual listeners, and thereby help them engage in new, more creative, immersive, and dynamic musical experiences. In the future, music experiences will demand far richer musical information, that supplements digital audio. FAST foresees that music content will be packaged in a flexible, structured way that combines audio recordings with rich, layered, standardized metadata to support interactive and adaptive musical experiences. The core unifying notion of FAST is the embodiment of these packages as digital music objects, constructed using the semantic web concepts of ontologies, linked data, and RDF. FAST therefore proposes to lay the foundations for a new generation of 'semantic audio' technologies that underpin diverse future music experiences. This keynote will describe the overall vision of FAST, and by highlighting some key outcomes (including some live demos), it will explore the notion of digital music objects and where they occur in the music production–consumption value chain.

4. Abstracts

4.1. P1. Poster Session 1

Session Chair: Hanna Järveläinen

P1.1. DAW-Integrated Beat Tracking for Music Production

Brett Dalton, David Johnson, and George Tzanetakis

Rhythm analysis is a well-researched area in music information retrieval that has many useful applications in music production. In particular, it can be used to synchronize the tempo of audio recordings with a digital audio workstation (DAW). Conventionally, this is done by stretching recordings over time, however, this can introduce artifacts and alter the rhythmic characteristics of the audio. Instead, this research explores how rhythm analysis can be used to do the reverse by synchronizing a DAW's tempo to a source recording. Drawing on research by Percival and Tzanetakis, a simple beat extraction algorithm was developed and integrated with the Renoise DAW. The results of this experiment show that, using user input from a DAW, even a simple algorithm can perform on par with popular packages for rhythm analysis, such as BeatRoot, IBT, and audio.

P1.2. Interaction-Based Analysis of Freely Improvised Music

Stefano Kalonaris

This paper proposes a computational method for the analysis and visualization of structure in freely improvised musical pieces, based on source separation and interaction patterns. A minimal set of descriptive axes is used for eliciting interaction modes, regions, and transitions. To this end, a suitable unsupervised segmentation model is selected based on the author's ground truth, and is used to compute and compare event boundaries of the individual audio sources. While still at a prototypal stage of development, this method offers useful insights for evaluating a musical expression that lacks

formal rules and protocols, including musical functions (e.g., accompaniment, solo, etc.) and form (e.g., verse, chorus, etc.).

P1.3. Mechanical Entanglement: A Collaborative Haptic-Music Performance

Alexandros Kontogeorgakopoulos, George Sioros, and Odysseas Klissouras

Mechanical Entanglement is a musical composition for three performers. Three force feedback devices each containing two haptic faders are mutually coupled using virtual linear springs and dampers. During the composition, the performers feel each other's gestures and collaboratively process the music material. The interaction's physical modelling parameters are modified during the different sections of the composition. An algorithm, which process three stereo channels, is stretching in and out-of-sync three copies of the same music. The performers are controlling the stretching algorithm and an amplitude modulation effect, both applied to recognizable classical and contemporary music recordings. Each of them is substantially modifying the length and the dynamics of the music and is simultaneously affecting subtly or abruptly the gestural behavior of the other performers. At fixed points during the composition, the music becomes gradually in sync and the performers realign their gestures. This phasing game between gestures and sound creates tension and emphasizes the physicality of the performance.

P1.4. State Dependency—Audiovisual Interaction through Brain States

Patrick Neff, Jan Schacher, and Daniel Bisig

Artistic installations using brain–computer interfaces (BCI) to interact with media in general, and sound specifically, have become increasingly numerous in the last years. Brain or mental states are commonly used to drive musical scores or sound generation as well as visuals. Closed loop setups can emerge here, which are comparable to the propositions of neurofeedback (NFB). The aim of our audiovisual installation, State Dependency, driven by brain states and motor imagery, was to enable the participant to engage in unbound exploration of movement through sound and space unmediated by one's corpo-reality. With the aid of an adaptive feedback loop, perception is taken to the edge. We deployed a BCI to collect motor imagery, and visual and cognitive neural activity to calculate the approximate entropy (a second order measure of neural signal activity), which was in turn used to interact with the surrounding Immersive Lab installation. The use of entropy measures on motor imagery and various sensory modalities generates a highly accessible, reactive, and immediate experience, transcending the common limitations of the BCI technology. State dependency goes beyond the common practice of abstract routing between mental or brain with external audiovisual states. It provides new territory of unrestrained kinesthetic and polymodal exploration in an immersive audiovisual environment.

P1.5. Perceptual Evaluation of Modal Synthesis for Impact-Based Sounds

Adrián Barahona and Sandra Pauletto

The use of real-time sound synthesis for sound effects can improve the sound design of interactive experiences, such as video games. However, synthesized sound effects can often be perceived as synthetic, which hampers their adoption. This paper aims to determine whether sounds synthesized using filter-based modal synthesis are perceptually comparable to sounds directly recorded. Sounds from four different materials that showed clear modes were recorded and synthesized using filter-based modal synthesis. Modes are the individual sinusoidal frequencies at which objects vibrate when excited. A listening test was conducted where participants were asked to identify, in isolation, whether a sample was recorded or synthesized. Results show that recorded and synthesized samples are indistinguishable from each other. The study outcome proves that, for the analyzed materials, filter-based modal synthesis is a suitable technique to synthesize hit sounds in real-time without perceptual compromises.

P1.6. VIBRA—Technical and Artistic Issues in an Interactive Dance Project

Andreas Bergsland, Sigurd Saue, and Pekka Stokke

The paper presents the interactive dance project, VIBRA, based on two workshops that took place in 2018. The paper presents the technical solutions applied and discusses artistic and expressive experiences. Central to the discussion is how the technical equipment, implementation, and mappings to different media affected the expressive and experiential reactions of the dancers.

P1.7. Musical Tempo and Key Estimation using Convolutional Neural Networks with Directional Filters

Hendrik Schreiber and Meinard Müller

In this article, we explore how the different semantics of spectrograms' time and frequency axes can be exploited for musical tempo and key estimation using convolutional neural networks (CNNs). By addressing both tasks with the same network architectures ranging from shallow, domain-specific approaches to deep variants with directional filters, we show that axis-aligned architectures perform similarly well as common VGG-style networks developed for computer vision, while being less vulnerable to confounding factors and requiring fewer model parameters.

P1.8. The Viking HRTF Dataset

Simone Spagnol, Kristján Bjarki Purkhús, Runar Unnthórsson, and Sverrir Karl Björnsson

This paper describes the Viking HRTF dataset, a collection of head related transfer functions (HRTFs) measured at the University of Iceland. The dataset includes fullsphere HRTFs measured on a dense spatial grid (1513 positions) with a KEMAR mannequin with 20 different artificial left pinnae attached, one at a time. The artificial pinnae were previously obtained through a custom molding procedure from 20 different lifelike human heads. The analyses of results reported here suggest that the collected acoustical measurements are robust, reproducible, and faithful to reference KEMAR HRTFs, and that the material hardness has a negligible impact on the measurements compared to the pinna shape. The purpose of the present collection, which is available for free download, is to provide accurate input data for future investigations on the relation between HRTFs and anthropometric data through machine learning techniques or other state-of-the-art methodologies.

P1.9. Performing with Sound Sample-Controlled Gloves and Light-Controlled Arms

Justin Pecquet, Fotis Moschos, David Fierro, and Frank Pecquet

Interacting with media: The TransTeamProject (T3P) works on developing interactive gloves techniques, and other materials, with sound and/or visual samples. Piamenca continues the work developed in Transpiano1 with a specific emphasis on visual content, such as transforming sound into lights, in this case, together with a strong vernacular inspiration (Flamenco). The T3P creative project is involved with art music—as opposed to commercial music—together with technical perspectives. After contextualizing the state-of-the-art in the specific field of "body gesture technology", this paper will explain how Piamenca relates to computers in a practical sense—methods and processes to produce media transformations (both audio and visual)—and will comment on their integration in terms of sound, music, and audio-visual performance. It will finally demonstrate some ideas, such as trans-music orientations, regarding enhancement theories in relation with the transhumanism movement.

P1.10. Melody Identification in Standard MIDI Files

Zheng Jiang and Roger Dannenberg

Melody identification is an important early step in music analysis. This paper presents a tool to identify the melody in each measure of a standard MIDI file. We also share an open dataset of manually labeled music for researchers. We used a Bayesian maximum-likelihood approach and dynamic programming as the basis of our work. We trained parameters on data sampled from the

million song dataset and tested on a dataset including 1703 measures of music from different genres. Our algorithm achieved an overall accuracy of 89% in the test dataset. We compare our results to previous work.

P1.11. Automatic Chord-Scale Recognition Using Harmonic Pitch Class Profiles

Emir Demirel, Baris Bozkurt, and Xavier Serra

This study focuses on the application of different computational methods to carry out a "modal harmonic analysis" for Jjzz improvisation performances by modeling the concept of chord-scales. The chord-scale theory is a theoretical concept that explains the relationship between the harmonic context of a musical piece and the possible scale types to be used for improvisation. This work proposes different computational approaches for the recognition of the chord-scale type in an improvised phrase given the harmonic context. We curated a dataset to evaluate different chord-scale recognition approaches proposed in this study, where the dataset consists of around 40 min of improvised monophonic jazz solo performances. The dataset is publicly available and shared on freesound.org. To achieve the task of chord-scale type recognition, we propose one rule-based, one probabilistic, and one supervised learning method. All proposed methods use harmonic pitch class profile (HPCP) features for classification. We observed an increase in the classification score when learned chord-scale models are filtered with predefined scale templates, indicating that the incorporation of prior domain knowledge into learned models is beneficial. This study has its novelty in presenting a first computational analysis on chord-scales in the context of jazz improvisation.

4.2. D1. Demo Session 1

Session Chair: Alberto Peinado

D1.1. Interacting with Digital Resonators by Acoustic Excitation

Max Neupert and Clemens Wegener

This demo presents an acoustic interface which allows direct excitation of digital resonators (digital waveguides, lumped models, modal synthesis, and sample convolution). Parameters are simultaneously controlled by the touch position on the same surface. The experience is an intimate and intuitive interaction with sound for percussive and melodic play.

D1.2. Melody Slot Machine

Masatoshi Hamanaka

This paper describes our interactive music system called the "Melody Slot Machine", which enables control of a holographic performer. Although many interactive music systems have been proposed, manipulating performances in real time is difficult for musical novices because melody manipulation requires expert knowledge. Therefore, we developed the Melody Slot Machine to provide an experience of manipulating melodies by enabling users to freely switch between two original melodies and morphing melodies.

D1.3. OM-AI: A Toolkit to Support AI-Based Computer-Assisted Composition Work in OpenMusic

Anders Vinjar and Jean Bresson

We present ongoing works exploring the use of artificial intelligence and machine learning in computer-assisted music composition. The OM-AI library for OpenMusic implements well-known techniques for data classification and prediction, in order to integrate them in composition workflows. We give examples using simple musical structures, highlighting possible extensions and applications.

D1.4. URALi: A Proposal of an Approach to Real-Time Audio Synthesis in Unity

Enrico Dorigatti

This paper aims to give a basic overview about the URALi (Unity Realtime Audio Library) project, that is currently under development. URALi is a library that aims to provide a collection of software tools to realize real-time sound synthesis in applications and software developed with Unity.

D1.5. A Sequencer with Decoupled Track Timing

Silvan David Peter and Gerhard Widmer

Sequencers almost exclusively share the trait of a single master clock.

Each track is laid out on an isochronously spaced sequence of beat positions. Vertically aligned positions are expected to be in synchrony as all tracks refer to the same clock. In this work, we present an experimental implementation of a decoupled sequencer with different underlying clocks. Each track is sequenced by the peaks of a designated oscillator. These oscillators are connected in a network and influence each other's periodicities. A familiar grid-type graphical user interface is used to place notes on beat positions of each of the interdependent but asynchronous tracks. Each track clock can be looped and node points specify the synchronization of multiple tracks by tying together specific beat positions. This setup enables simple global control of microtiming and polyrhythmic patterns.

D1.6. Musicypher: Music for Message Encryption

Víctor Jaime Marín and Alberto Peinado

An Android application has been developed to encrypt messages using musical notes that can be automatically played from the smartphone and/or stored in a midi file to be transmitted over any available connection. The app has been designed to recover the original message by on-the-fly detection of the notes played by a different device. The main objective of this project is to make known the relationship between cryptography and music showing old systems (XVII century) implemented in modern devices, the smartphones, using the tools they provide us, such as the microphone, speakers, and internal storage.

D1.7. A Platform for Processing Sheet Music and Developing Multimedia Application

Fu-Hai Frank Wu

Imagine that when reading sheet music on computing devices, users could listen to audio that synchronizes with the sheet. To this end, the sheet music must be acquired, analyzed, and transformed into digitized information of the melody, rhythm, duration, chord, expressiveness, and physical location of scores. As we know, optical music recognition (OMR) is an appropriate technology to approach this purpose. However, the commercial OMR system of numbered music notation is not available to the best as our knowledge. In the paper, we demonstrate our proprietary OMR system and show three human-interactive applications: Sheet music browser and multimodal accompanying and games for sight-reading of sheet music. With the illustration, we hope to foster usage and obtain valuable opinions of the OMR system and its applications.

D1.8. Capturing the Reaction Time to Distinguish between Voice and Music

Alejandro Villena-Rodríguez, Lorenzo J. Tardón, Isabel Barbancho, Ana M. Barbancho, Irene Gómez-Plazas, and María-José Varela-Salinas

Reaction times (RTs) are an important source of information in experimental psychology and EEG data analysis. While simple auditory RTs have been widely studied, response times when discriminating between two different auditory stimuli have not been determined yet. The purpose of this experiment is to measure the RT for the discrimination between two different auditory stimuli: Speech and instrumental music.

D1.9. Physical Models and Real-Time Control with the Sensel Morph

Silvin Willemsen, Stefan Bilbao, Nikolaj Andersson, and Stefania Seran

In this demonstration, we present novel physical models controlled by the Sensel Morph interface.

4.3. S1. Oral Session 1: Sonic Interactions

Session Chair: Stefania Serafin

S1.1. Towards a High-Performance Platform for Sonic Interaction Interfaces

Stefano Fasciani and Manohar Vohra

In this paper, we introduce a hardware platform to prototype interfaces of demanding sonic interactive systems. We target applications featuring a large array of analog sensors requiring data acquisition and transmission to computers at fast rates, with a low latency and high band- width. This work is part of an ongoing project which aims to provide designers with a cost effective and accessible platform for fast prototyping of complex interfaces for sonic interactive systems or musical instruments. High performances are guaranteed by an SoC FPGA. The functionality of the platform can be customized without requiring significant technical expertise. In this paper, we discuss the principles, current design, and preliminary evaluation against common microcontroller-based platforms. The proposed platform can sample up to 96 analog channels at rates of up to 24 kHz and stream the data via UDP to computers with a sub millisecond latency.

S1.2. Digital Manufacturing for Musical Applications: A Survey of the Current Status and Future Outlook

Doga Cavdir

In the design of new musical instruments, from acoustic to digital, merging conventional methods with new technologies has been one of the most commonly adopted approaches. The incorporation of prior design expertise with experimental or sometimes industrial methods suggests new directions in both the design for musical expression and the development of new manufacturing tools. This paper describes key concepts of the digital manufacturing processes in musical instrument design. It provides a review of current manufacturing techniques which are commonly used to create new musical interfaces and discusses future directions of digital fabrication which are applicable to numerous areas in music research, such as digital musical instrument (DMI) design, interaction design, acoustics, performance studies, and education. Additionally, the increasing availability of digital manufacturing tools and fabrication labs all around the world make these processes an integral part of the design and music classes. Examples of digital fabrication labs and manufacturing techniques used in education for student groups whose ages range from elementary to university level are presented. In the context of this paper, it is important to consider how the growing fabrication technology will influence the design and fabrication of musical instruments, as well as what forms of new interaction methods and aesthetics might emerge.

S1.3. Real Time Audio Digital Signal Processing with Faust and the Teensy

Romain Michon, Yann Orlarey, Stéphane Letz, and Dominique Fober

In the design of new musical instruments, from acoustic to digital, merging conventional methods with new technologies has been one of the most commonly adopted approaches. Incorporation of prior design expertise with experimental or sometimes industrial methods suggests new directions in both the design for musical expression and the development of new manufacturing tools. This paper describes key concepts of digital manufacturing processes in musical instrument design. It provides a review of the current manufacturing techniques which are commonly used to create new musical interfaces and discusses future directions of digital fabrication which are applicable to numerous areas

in music research, such as digital musical instrument (DMI) design, interaction design, acoustics, performance studies, and education. Additionally, the increasing availability of digital manufacturing tools and fabrication labs all around the world make these processes an integral part of the design and music classes. Examples of digital fabrication labs and manufacturing techniques used in education for student groups whose ages range from elementary to university level are presented. In the context of this paper, it is important to consider how the growing fabrication technology will influence the design and fabrication of musical instruments, as well as what forms of new interaction methods and aesthetics might emerge.

S1.4. Sound Design through Large Audience Interaction

Kjetil Falkenberg Hansen, Martin Ljungdahl-Eriksson, and Ricardo Atienza

In collaboration with Volvo Cars, we presented a novel design tool to a large public of approximately three million people at the three leading motor shows in 2017 in Geneva, Shanghai, and New York. The purpose of the tool was to explore the relevance of interactive audio-visual strategies for supporting the development of sound environments in future silent cars, i.e., a customized sonic identity that would alter the sonic ambience for the driver and by-passers. This new tool should be able to efficiently collect non-experts' sonic preferences for different given contexts. The design process should allow for a high-level control of complex synthesized sounds. The audience interacted individually using a single touch selection of color from five palettes and applied it by pointing to areas in a color-book painting showing a road scene. Each palette corresponded to a sound, and the color nuance in the palette corresponded to certain tweaking of the sound. In effect, the user selected and altered each sound, added it to the composition, and finally would hear a mix of layered sounds based on the coloring of the scene. The installation involved large touch screens with high quality headphones. In the study presented here, we examine differences in sound preferences between two audiences and a control group, and evaluate the feasibility of the tool based on the sound designs that emerged.

S1.5. Evaluating a Continuous Sonic Interaction: Comparing a Performable Acoustic and Digital Everyday Sound

Fiona Keenan and Sandra Pauletto

This paper reports on the procedure and results of an experiment to evaluate a continuous sonic interaction with an everyday wind-like sound created by both acoustic and digital means. The interaction is facilitated by a mechanical theatre sound effect, an acoustic wind machine, which is performed by participants. This work is part of wider research into the potential of theatre sound effect designs as a means to study multisensory feedback and continuous sonic interactions. An acoustic wind machine is a mechanical device that affords a simple rotational gesture to a performer; turning its crank handle at varying speeds produces a wind-like sound. A prototype digital model of a working acoustic wind machine is programmed, and the acoustic interface drives the digital model in performance, preserving the same tactile and kinesthetic feedback across the continuous sonic interactions. Participants' performances are elicited with sound stimuli produced from simple gestural performances of the wind-like sounds. The results of this study show that the acoustic wind machine is rated as significantly easier to play than its digital counterpart. Acoustical analysis of the corpus of participants' performances suggests that the mechanism of the wind machine interface may play a role in guiding their rotational gestures.

4.4. S2. Oral Session 2: Nordic SMC

Session Chair: Vesa Välimäki

S2.1. Adaptive Loudness Compensation in Music Listening

Leonardo Fierro, Jussi Rämö, and Vesa Välimäki

The need for loudness compensation is a well-known fact arising from the nonlinear behavior of human sound perception. Music and other sounds are mixed and mastered at a certain loudness level, usually louder than the level at which they are commonly played. This implies a change in the perceived spectral balance of the sound, which is largest in the low frequency range. As the volume setting in music playing is decreased, a loudness compensation filter can be used to boost the bass appropriately, so that the low frequencies are still heard well and the perceived spectral balance is preserved. The present paper proposes a loudness compensation function derived from the standard equal-loudness-level contours and its implementation via a digital first-order shelving filter. Results of a formal listening test validate the accuracy of the proposed method.

S2.2. Toward Automatic Tuning of the Piano

Joonas Tuovinen, Jamin Hu, and Vesa Välimäki

The tuning of a piano is a complicated and time-consuming process, which is usually left for a professional tuner. To make the process faster and non-dependent on the skills of a professional tuner, a semi-automatic piano tuning system is developed. The aim of the system is to help a nonprofessional person tune a grand piano with the help of a computer and a motorized tuning machine. The system is composed of an aluminum frame, a stepper motor, an Arduino processor, a microphone, and a laptop computer. The stepper motor changes the tuning of the piano strings by turning the pins connected to them whereas the aluminum frame holds the motor in place. The Arduino controls the motor. The microphone and the computer are used as a part of a closed loop control system, which is used to tune the strings automatically. The control system tunes the strings by minimizing the difference between the current and optimal fundamental frequency. The current fundamental frequency is obtained with an inharmonicity coefficient estimation algorithm, and the optimal fundamental frequency is calculated with a novel tuning process, called the connected reference interval (CRI) tuning. With the CRI process, a tuning close to that of a professional tuner is achieved with a deviation of 2.5 cents (RMS) between the keys, A0 and G5, and 8.1 cents (RMS) between G#5 and C8, where the tuner's results are not very consistent.

S2.3. Real-Time Control of Large-Scale Modular Physical Models Using the Sensel Morph

Silvin Willemsen, Nikolaj Andersson, Stefania Seran, and Stefan Bilbao

In this paper, the implementation, instrument design, and control issues surrounding a modular physical modelling synthesis environment are described. The environment is constructed as a network of strings and a resonant plate, accompanied by user-defined connections and excitation models. The bow, in particular, is a novel feature in this setting. The system as a whole is simulated using finite difference (FD) methods. The mathematical formulation of these models is presented, alongside several new instrument designs, together with a real-time implementation in JUCE using FD methods. Control is through the Sensel Morph.

S2.4. An Interactive Music Synthesizer for Gait Training in Neurorehabilitation

Prithvi Kantan and Soa Dahl

Rhythm-based auditory cues have been shown to significantly improve walking performance in patients with numerous neurological conditions. This paper presents the design, implementation, and evaluation of a gait training device capable of real-time synthesis and automated manipulation

of rhythmic musical stimuli, as well as auditory feedback based on measured walking parameters. The proof-of-concept was evaluated with six healthy participants, as well as through critical review by one neurorehabilitation specialist. Stylistically, the synthesized music was found by participants to be conducive to movement, but not uniformly enjoyable. The gait capture/feedback mechanisms functioned as intended, although discrepancies between measured and reference gait parameter values may necessitate a more robust measurement system. The specialist acknowledged the potential of the gait measurement and auditory feedback as novel rehabilitation aids, but stressed the need for additional gait measurements, superior feedback responsiveness, and greater functional versatility in order to cater to individual patient needs. Further research must address these findings, and tests must be conducted on real patients to ascertain the utility of such a device in the field of neurorehabilitation.

S2.5. From Vocal Sketching to Sound Models by Means of a Sound-Based Musical Transcription System

Claudio Panariello, Mattias Sköld, Emma Frid, and Roberto Bresin

This paper explores how notation developed for the representation of sound-based musical structures could be used for the transcription of vocal sketches representing expressive robot movements. A mime actor initially produced expressive movements which were translated to a humanoid robot. The same actor was then asked to illustrate these movements using vocal sketching. The vocal sketches were transcribed by two composers using sound-based notation. The same composers later synthesized new sonic sketches from the annotated data. Different transcriptions and synthesized versions of these were compared in order to investigate how the audible outcome changes for different transcriptions and synthesis routines. This method provides a palette of sound models suitable for the sonification of expressive body movements.

S2.6. Tempo and Metrical Analysis by Tracking Multiple Metrical Levels Using Autocorrelation

Olivier Lartillot and Didier Grandjean

We present a method for tempo estimation from audio recordings based on signal processing and peak tracking, and not depending on training on ground-truth data. First, an accentuation curve, emphasizing the temporal location and accentuation of notes, is based on the detection of bursts of energy localized in time and frequency. This enables the detection of notes in a dense polyphonic texture, while ignoring spectral fluctuation produced by vibrato and tremolo. Periodicities in the accentuation curve are detected using an improved version of the autocorrelation function. Hierarchical metrical structures, composed of a large set of periodicities in pairwise harmonic relationships, are tracked over time. In this way, the metrical structure can be tracked even if the rhythmical emphasis switches from one metrical level to another. This approach, compared to all the other participants to the MIREX Audio Tempo Extraction from 2006 to 2018, is the third best one among those that can track tempo variations. While the two best methods are based on machine learning, our method suggests a way to track tempo founded on signal processing and heuristics-based peak tracking. Besides, the approach offers for the first time a detailed representation of the dynamic evolution of the metrical structure. The method is integrated into MIRtoolbox, a Matlab toolbox that is freely available.

4.5. S3. Oral Session 3: Augmented and Virtual Realities

Session Chair: Marcella Mandanici

S3.1. Comparison and Implementation of Data Transmission Techniques through Analog Audio Signals in the Context of Augmented Mobile Instruments

Romain Michon, Yann Orlarey, Stéphane Letz, and Dominique Fober

Augmented mobile instruments combine digitally-fabricated elements, sensors, and smartphones to create novel musical instruments. Communication between the sensors and the smartphone can be challenging as a universal lightweight way to connect external elements to this type of device does

not exist. In this paper, we investigate the use of two techniques to transmit sensor data through the built-in audio jack input of a smartphone: Digital data transmission using the Bell 202 signaling technique, and analog signal transmission using digital amplitude modulation and demodulation with Goertzel filters. We also introduce tools to implement such systems using the FAUST programming language and the Teensy development board.

S3.2. Mass-Interaction Physical Models for Sound and Multi-Sensory Creation: Starting Anew

Jerome Villeneuve and James Leonard

Mass-interaction methods for sound synthesis, and more generally for digital artistic creation, have been studied and explored for over three decades by a multitude of researchers and artists. However, for a number of reasons, this research has remained rather confidential, subsequently overlooked, and often considered as the odd-one-out of physically based synthesis methods, of which many have grown exponentially in popularity over the last 10 years. In the context of a renewed research effort led by the authors on this topic, this paper aims to reposition mass interaction physical modelling in the contemporary fields of sound and music computing and digital arts: What are the core concepts? The end goals? And, more importantly, which relevant perspectives can be foreseen in this current day and age? Backed by recent developments and experimental results, including 3D mass-interaction modelling and emerging non-linear effects, this proposed reflection casts a first canvas for an active, and resolutely outreaching, research on mass-interaction physical modelling for the arts.

S3.3. Exploring the Effects of Diegetic and Non-Diegetic Audiovisual Cues on Decision-Making in a Virtual Reality

Anıl Çamcı

The user experience of a virtual reality intrinsically depends upon how the underlying system relays information to the user. Auditory and visual cues that make up the user interface of a VR help users make decisions on how to proceed in a virtual scenario. These interfaces can be diegetic (i.e., presented as part of the VR) or non-diegetic (i.e., presented as an external layer superimposed onto the VR). In this paper, we explore how auditory and visual cues of diegetic and non-diegetic origins affect a user's decision making process in a VR. We present the results of a pilot study, where users are placed into virtual situations and are expected to make choices upon conflicting suggestions as to how to complete a given task. We analyze the quantitative data pertaining to user preferences for modality and diegetic-quality. We also discuss the narrative effects of the cue types based on a follow-up survey conducted with the users.

S3.4. OSC-XR: A Toolkit for Extended Reality Immersive Music Interfaces

David Johnson, Daniela Damian, and George Tzanetakis

Currently, developing immersive music environments for extended reality (XR) can be a tedious process requiring designers to build 3D audio controllers from scratch. OSCXR is a toolkit for Unity intended to speed up this process through rapid prototyping, enabling research in this emerging field. Designed with multi-touch OSC controllers in mind, OSC-XR simplifies the process of designing immersive music environments by providing prebuilt OSC controllers and Unity scripts for designing custom ones. In this work, we describe the toolkit's infrastructure and perform an evaluation of the controllers to validate the generated control data. In addition to OSC-XR, we present UnityOscLib, a simplified OSC library for Unity utilized by OSC-XR. We implemented three use cases, using OSCXR, to inform its design and demonstrate its capabilities. The Sonic Playground is an immersive environment for controlling audio patches. Hyperemin is an XR hyper instrument environment in which we augment a physical theremin with OSC-XR controllers for real-time control of audio processing. Lastly, we add OSC-XR controllers to an immersive T-SNE visualization of music genre data for enhanced exploration

and sonification of the data. Through these use cases, we explore and discuss the affordances of OSC-XR and immersive music interfaces.

S3.5. No Strings Attached: Force and Vibrotactile Feedback in a Guitar Simulation

Andrea Passalenti, Razvan Paisa, Niels Christian Nilsson, Nikolaj S. Andersson, Federico Fontana, Rolf Nordahl, and Stefania Seran

In this paper, we propose a multisensory simulation of plucking guitar strings in virtual reality. The auditory feedback is generated by a physics-based simulation of guitar strings, and haptic feedback is provided by a combination of high fidelity vibrotactile actuators and a Phantom Omni haptic device. Moreover, we present a user study (n = 29) exploring the perceived realism of the simulation and the relative importance of force and vibrotactile feedback for the creation of a realistic experience of plucking virtual strings. The study compares four conditions: No haptic feedback, vibrotactile feedback, force feedback, and a combination of force and vibrotactile feedback. The results indicate that the combination of vibrotactile and force feedback elicits the most realistic experience, and during this condition, the participants were less likely to inadvertently hit strings after the intended string had been plucked. Notably, no statistically significant differences were found between the conditions involving either vibrotactile or force feedback, which points towards an indication that haptic feedback is important but does not need to be of high fidelity in order to enhance the quality of the experience.

4.6. P2. Poster Session 2

Session Chair: Anja Volk

P2.1. RaveForce: A Deep Reinforcement Learning Environment for Music Generation

Qichao Lan, Jim Tørresen, and Alexander Refsum Jensenius

RaveForce is a programming framework designed for a computational music generation method that involves audio sample level evaluation in symbolic music representation generation. It comprises a Python module and a SuperCollider quark. When connected with deep learning frameworks in Python, RaveForce can send the symbolic music representation generated by the neural network as Open Sound Control messages to the SuperCollider for non-real-time synthesis. The SuperCollider can convert the symbolic representation into an audio file which will be sent back to the Python as the input of the neural network. With this iterative training, the neural network can be improved with deep reinforcement learning algorithms, taking the quantitative evaluation of the audio file as the reward. In this paper, we find that the proposed method can be used to search new synthesis parameters for a specific timbre of an electronic music note or loop.

P2.2. Music Temperaments Evaluation Based on Triads

Tong Meihui and Satoshi Tojo

It is impossible for one temperament to achieve optimally both consonance and modulation. The dissonance level has been calculated by the ratio of two pitch frequencies; however, in the current homophonic music, the level should be measured by chords, especially by triads. In this research, we propose to quantify them as dissonance index of triads (DIT). We select eight well-known temperaments and calculate seven diatonic chords in 12 keys and compare the weighted average and standard deviation to quantify the consonance, and then we visualize our experimental results in a two-dimensional chart to compare the tradeoffs between consonance and modulation.

P2.3. Composing Space in the Space: An Augmented and Virtual Reality Sound Spatialization System

Giovanni Santini

This paper describes a tool for gesture-based control of sound spatialization in augmented and virtual reality (AR and VR). While the increased precision and availability of sensors of any kind has made possible, in the last 20 years, the development of a considerable number of interfaces for sound spatialization control through gesture, their integration with VR and AR has not been fully explored

yet. Such technologies provide an unprecedented level of interaction, immersivity, and ease of use, by letting the user visualize and modify the position, trajectory, and behavior of sound sources in 3D space. Like VR/AR painting programs, the application allows the drawing of lines that have the function of 3D automations for spatial motion. The system also stores information about the movement speed and directionality of the sound source. Additionally, other parameters can be controlled from a virtual menu. The possibility to alternate AR and VR allows switching between different environments (the actual space where the system is located or a virtual one). Virtual places can also be connected to different room parameters inside the spatialization algorithm.

P2.4. Graph Based Physical Models for Sound Synthesis

Pelle Juul Christensen and Stefania Seran

We focus on physical models in which multiple strings are connected via junctions to form graphs. Starting with the case of the 1D wave equation, we show how to extend it to a string branching into two other strings, and from there how to build complex cyclic and acyclic graphs. We introduce the concept of dense models and show that a discretization of the 2D wave equation can be built using our methods, and that there are more efficient ways of modelling 2D wave propagation than a rectangular grid. We discuss how to apply Dirichlet and Neumann boundary conditions to a graph model, and show how to compute the frequency content of a graph using common methods. We then prove general lower and upper bounds of computational complexity. Lastly, we show how to extend our results to other kinds of acoustical objects, such as linear bars, and how to add dampening to a graph model. A reference implementation in MATLAB and an interactive JUCE/C++ application is available online.

P2.5. ADPET: Exploring the Design, Pedagogy, and Analysis of a Mixed Reality Application for Piano Training

Lynda Gerry, Soa Dahl, and Stefania Seran

One of the biggest challenges in learning how to play a musical instrument is learning how to move one's body with a nuanced physicality. Technology can expand the available forms of physical interactions to help cue specific movements and postures. This cueing can reinforce new sensorimotor couplings to enhance motor learning and performance. Using mixed reality (MR), we present a system that allows students to share a first-person audiovisual perspective with a piano teacher. Students place their hands into the virtual gloves of a teacher. Motor learning and audio-motor associations are reinforced through motion feedback and spatialized audio. The Augmented Design to Embody a Piano Teacher (ADEPT) application is an early design prototype of this piano training system.

P2.6. A Model Comparison for Chord Prediction on the Annotated Beethoven Corpus

Kristoer Landsnes, Liana Mehrabyan, Victor Wiklund, Robert Lieck, Fabian Moss, and Martin Rohrmeier

This paper models the predictive processing of chords using a corpus of Ludwig van Beethoven's string quartets. A recently published dataset consisting of expert harmonic analyses of all Beethoven string quartets was used to evaluate an n-gram language model as well as a recurrent neural network (RNN) architecture based on long-short-term memory (LSTM). We compare model performances over different periods of Beethoven's creative activity and provide a baseline for future research on the predictive processing of chords in full Roman numeral representation on this dataset.

P2.7. Sonic Characteristics of Robots in Films

Adrian B. Latupeirissa, Emma Frid, and Roberto Bresin

Robots are increasingly becoming an integral part of our everyday life. Expectations on robots could be influenced by how robots are represented in science fiction films. We hypothesize that sonic interaction design for real-world robots may find inspiration from the sound design of fictional robots. In this paper, we present an exploratory study focusing on the sonic characteristics of robot sounds

in films. We believe that the findings from the current study could be of relevance for future robotic applications involving the communication of internal states through sounds, as well for sonification of expressive robot movements. Excerpts from five films were annotated and analyzed using long time average spectrum (LTAS). As an overall observation, we found that a robot's sonic presence is highly related to the physical appearance of robots. Preliminary results show that most of the robots analyzed in this study have "metallic" voice qualities, matching the material of their physical form. Characteristics of robot voices show significant differences compared to the voices of human characters; the fundamental frequency of robotic voices is either shifted to higher or lower values, and the voices span over a broader frequency band.

P2.8. Virtual Reality Music Intervention to Reduce Social Anxiety in Adolescents Diagnosed with Autism Spectrum Disorder

Ali Adjorlu, Nathaly Belen Betancourt Barriga, and Stefania Seran

This project investigates the potentials of head-mounted-display (HMD)-based virtual reality (VR) that incorporates musical elements as a tool to perform exposure therapy. This is designed to help adolescents diagnosed with autism spectrum disorder (ASD) to deal with their social anxiety. An application was developed that combines the possibility of singing in VR while a virtual audience provides feedback. A pilot test was conducted on four adolescents diagnosed with ASD from a school for adolescents with special needs in Denmark. All four participants had shown signs of social anxiety according to their teachers. The initial results from this pilot study indicate that despite the participants' ASD, they were capable of singing in front of the virtual audience without reporting a major level of social anxiety.

P2.9. Teach Me Drums: Learning Rhythms through the Embodiment of a Drumming Teacher in Virtual Reality

Mie Moth-Poulsen, Tomasz Bednarz, Volker Kuchelmeister, and Stefania Seran

This paper investigates how to design an embodied learning experience of a drumming teacher playing hand drums, to aid higher rhythm understanding and accuracy. By providing novices the first-person perspective of a drumming teacher while learning to play a West-African djembe drum, participants' learning was measured objectively by their ability to follow the drumming teacher's rhythms.

Participants' subjective learning was assessed through a self-assessment questionnaire measuring aspects of flow, user-experience, oneness, and presence. Two test iterations were conducted. In both, no significance difference was found in participants' ability to follow the drumming teacher's tempo for the experimental group exposed to the first-person perspective of the teacher in a virtual reality (VR) drum lesson versus the control group exposed to a 2D version of the stereoscopic drum lesson. A significant difference was found in the experimental group's presence scores in the first test iteration, and a significant difference was found in the experimental group's oneness scores in the second test iteration. Participants' subjective feelings indicated enjoyment and motivation to the presented learning technique in both groups.

P2.10. Real-Time Mapping of Periodic Dance Movements to Control Tempo in Electronic Dance Music

Lilian Jap and Andre Holzapfel

Dancing in beat to the music of one's favorite DJ oftentimes leads to a powerful and euphoric experience. In this study, we investigate the effect of putting a dancer in control of the music playback tempo based on a real-time estimation of the body rhythm and tempo manipulation of the audio. A prototype was developed and tested in collaboration with users, followed by a main study where the final prototype was evaluated. A questionnaire was provided to obtain ratings regarding the subjective experience, and open-ended questions were posed in order to obtain further insights for

future development. Our results imply the potential for enhanced engagement and enjoyment of the music when being able to manipulate the tempo, and document important design aspects for real-time tempo control.

P2.11. Increasing Access to Music in SEN Settings

Tom Davis, Daniel Pierson, and Ann Bevan

This paper presents some of the outcomes of a one year Higher Education Innovation Fund1 funded project examining the use of music technology to increase access to music for children within special educational need (SEN) settings. Despite the widely acknowledged benefits of interacting with music for children with SENs, there are a number of well documented barriers to access. These barriers take a number of forms, including financial, knowledge-based, or attitudinal. The aims of this project were to assess the current music technology provision in SEN schools within a particular part of the Dorset region, UK, to determine the barriers they were facing and develop strategies to help the schools overcome these barriers. An overriding concern for this project was to leave the schools with lasting benefit and meaningful change. As such an action research methodology was followed, which has at its heart an understanding of the participants as co-researchers, helping to ensure that any solutions presented met the needs of the stakeholders. The presumption by the researchers was that the schools needed new technology to help overcome barriers. However, although technological solutions to problems were presented to the school, it was found that the main issues were around the flexibility of equipment to be used in different locations, staff time, and staff attitudes to technology. These issues were addressed through the action research methodology to ensure that the technology designed worked for these particular use case scenarios.

4.7. D2. Demo Session 2

Session Chair: Hendrik Schreiber

D2.1. Interacting with Musebots (That Don't Really Listen)

Arne Eigenfeldt

TinySounds is a collaborative work for a live performer and musebot ensemble. Musebots are autonomous musical agents that interact, via messaging, to create a musical performance with or without human interaction.

D2.2. Extending Jamsketch: An Improvisation Support System

Akane Yasuhara, Junko Fujii, and Tetsuro Kitahara

We previously introduced JamSketch, a system which enabled users to improvise music by drawing a melodic outline. However, users could not control the rhythm and intensity of the generated melody. Here, we present extensions to JamSketch to enable rhythm and intensity control.

D2.3. Visualizing Music Genres Using a Topic Model

Swaroop Panda, Vinay P. Namboodiri, and Shatarupa Thakurta Roy

Music genres serve as an important meta-data in the field of music information retrieval and have been widely used for music classification and analysis tasks. Visualizing these music genres can thus be helpful for music exploration, archival, and recommendation. Probabilistic topic models have been very successful in modelling text documents. In this work, we visualize music genres using a probabilistic topic model. Unlike text documents, audio is continuous and needs to be sliced into smaller segments. We use simple MFCC features of these segments as musical words. We apply the topic model on the corpus and subsequently use the genre annotations of the data to interpret and visualize the latent space.

D2.4. CompoVOX: Real-Time Sonication of Voice

Daniel Hernán Molina Villota, Isabel Barbancho, and Antonio Jurado-Navas

An interactive application has been developed that allows sonification of the human voice and visualization of a graphic interface in relation to the sounds produced. This program has been developed in MAX MSP, and it takes the spoken voice signal, and from its treatment, it allows the generation of an automatic and tonal musical composition.

D2.5. Facial Activity Detection to Monitor Attention and Fatigue

Oscar Cobos, Jorge Munilla, Ana M. Barbancho, Isabel Barbancho, and Lorenzo J. Tardón

In this contribution, we present a facial activity detection system using image processing and machine learning techniques. Facial activity detection allows the monitoring of people's emotional states, attention, fatigue, reactions to different situations, etc. in a non-intrusive way. The designed system can be used in many fields, such as education and musical perception. Monitoring the facial activity of a person can help us to know if it is necessary to take a break, change the type of music that is being listened to, or modify the way of teaching the class.

D2.6. The Chordinator: An Interactive Music Learning Device

Eamon McCoy, John Greene, Jared Henson, James Pinder, Jonathon Brown, and Claire Arthur

The Chordinator is an interactive and educational music device consisting of a physical board housing a "chord stacking" grid. There is an 8 × 4 grid on the board which steps through each of the eight columns from left to right at a specified tempo, playing the chords you have built in each column. To build a chord, you place blocks on the board which represent major or minor thirds above blocks that designate a root (or bass) note represented as a scale degree. In the bottom row, the user specifies a bass (root) note, and any third blocks placed above it will add that interval above the bass note. Any third blocks placed above other third blocks add an additional interval above the prior one, creating a chord. There are three rows above each root, allowing either triads or seventh chords to be built. This interface combined with the board design is intended to create a simple representation of chord structure. Using the blocks, the user can physically "build" a chord using the most fundamental skills, in this case "stacking your thirds". One also learns which chords work the best in a sequence. It provides quick satisfaction and a fun, interactive way to learn about the structure of chords, and can even spark creativity as people build interesting progressions or try to recreate progressions they love from their favorite music.

D2.7. Automatic Chord Recognition in Music Education Applications

Sascha Grollmisch and Estefania Cano

In this work, we demonstrate the market-readiness of a recently published state-of-the-art chord recognition method, where automatic chord recognition is extended beyond major and minor chords to the extraction of seventh chords. To do so, the proposed chord recognition method was integrated in the Songs2See Editor, which already includes the automatic extraction of the main melody, bass line, beat grid, key, and chords for any musical recording.

D2.8. Sonic Sweetener Mug

Signe Lund Mathiesen, Derek Victor Byrne, and Qian Janice Wang

Eating is one of the most sensory of all activities that we take part in. Apart from tasting, it involves both the food and the environment. The multitude of different sensory inputs (from the smell of the food and the color of the plate, to the lighting in the room and the ambient soundscape) all affect the way we think about and perceive our food. Much like eating, listening is a fundamental part of most lives; and similar to the role of food, music can modulate our feelings, our mood, and our experiences

in life. This demo explores the common link between these two phenomena, specifically the way in which what we taste can be influenced by what we listen to.

4.8. S4. Oral Session 4: SMC Tools and Methodologies

Session Chair: Emma Frid

S4.1. A Framework for the Development and Evaluation of Graphical Interpolation for Synthesizer Parameter Mappings

Darrell Gibson and Richard Polfreman

This paper presents a framework that supports the development and evaluation of graphical interpolated parameter mapping for the purpose of sound design. These systems present the user with a graphical pane, usually two-dimensional, where synthesizer presets can be located. Moving an interpolation point cursor within the pane will then create new sounds by calculating new parameter values, based on the cursor position and the interpolation model used. The exploratory nature of these systems lends itself to sound design applications, which also have a highly exploratory character. However, populating the interpolation space with "known" preset sounds allows the parameter space to be constrained, reducing the design complexity otherwise associated with synthesizer-based sound design. An analysis of previous graphical interpolators is presented and from this a framework is formalized and tested to show its suitability for the evaluation of such systems. The framework has then been used to compare the functionality of a number of systems that have been previously implemented. This has led to a better understanding of the different sonic outputs that each can produce and highlighted areas for further investigation.

S4.2. Composing with Sounds: Designing an Object Oriented DAW for the Teaching of Sound-Based Composition

Stephen Pearse, Leigh Landy, Duncan Chapman, David Holland, and Mihai Eni

This paper presents and discusses the Compose with Sounds (CwS) Digital Audio Workstation (DAW) and its approach to sequencing musical materials. The system is designed to facilitate the composition within the realm of sound-based music wherein sound objects (real or synthesized) are the main musical unit of construction over traditional musical notes. Unlike traditional DAWs or graphical audio programming environments (such as Pure Data, Max MSP, etc.) that are based around interactions with sonic materials within tracks or audio graphs, the implementation presented here is based solely around sound objects. To achieve this, a bespoke cross-platform audio engine known as FSOM (Free Sound Object Mixer) was created in C++. To enhance the learning experience, imagery, dynamic 3D animations, and models are used to allow for efficient exploration and learning. All tools within the system are controlled by a flexible permissions system that allows users or workshop leaders to create sessions with specific features based on their requirements. The system is part of a suite of pedagogical tools currently in development for the creation of experimental electronic music.

S4.3. Insights in Habits and Attitudes Regarding Programming Sound Synthesizers: A Quantitative Study

Gordan Kreković

Sound synthesis represents an indispensable tool for modern composers and performers, but achieving desired sonic results often requires a tedious manipulation of various numeric parameters. In order to facilitate this process, a number of possible approaches have been proposed, but without a systematic user research that could help researchers to articulate the problem and to make informed design decisions. The purpose of this study is to fill that gap and to investigate attitudes and habits of sound synthesizer users. The research was based on a questionnaire answered by 122 participants, which, beside the main questions about habits and attitudes, covered questions about

their demographics, profession, educational background, and experience in using sound synthesizers. The results were quantitatively analyzed in order to explore relations between all those dimensions. The main results suggest that the participants more often modify or create programs than use existing presets or programs and that such habits do not depend on the participants' education, profession, or experience.

4.9. S5. Oral Session 5: Sound Synthesis and Analysis

Session Chair: Federico Avanzini

S5.1. Experimental Verification of Dispersive Wave Propagation on Guitar Strings

Dmitri Kartofelev, Joann Gustav Arro, and Vesa Välimäki

Experimental research into the fundamental acoustic aspects of musical instruments and other sound generating devices is an important part of the history of musical acoustics and of physics in general. This paper presented experimental proof of dispersive wave propagation on metal guitar strings. The high resolution experimental data of string displacement are gathered using video-kymographic high-speed imaging of the vibrating string. The experimental data are indirectly compared against a dispersive Euler–Bernoulli type model described by a PDE. In order to detect the minor wave features associated with the dispersion and distinguish them from other effects present, such as frequency-dependent dissipation, a second model lacking the dispersive (stiffness) term is used. Unsurprisingly, the dispersive effects are shown to be minor but definitively present. The results and methods presented here in general should find application in string instrument acoustics.

S5.2. Real-Time Modeling of Audio Distortion Circuits with Deep Learning

Eero-Pekka Damskägg, Lauri Juvela, and Vesa Välimäki

This paper studies deep neural networks for the modeling of audio distortion circuits. The selected approach is black-box modeling, which estimates model parameters based on the measured input and output signals of the device. Three common audio distortion pedals having a different circuit configuration and their own distinctive sonic character have been chosen for this study: The Ibanez Tube Screamer, the Boss DS-1, and the Electro-Harmonix Big Mu Pi. A feedforward deep neural network, which is a variant of the WaveNet architecture, is proposed for modeling these devices. The size of the receptive field of the neural network is selected based on the measured impulse–response length of the circuits. A real-time implementation of the deep neural network is presented, and it is shown that the trained models can be run in real-time on a modern desktop computer. Furthermore, it is shown that three minutes of audio is a sufficient amount of data for training the models. The deep neural network studied in this work is useful for real-time virtual analog modeling of nonlinear audio circuits.

S5.3. MI-GEN~: An Efficient and Accessible Mass-Interaction Sound Synthesis Toolbox

James Leonard and Jerome Villeneuve

Physical modelling techniques are now an essential part of digital sound synthesis, allowing for the creation of complex timbres through the simulation of virtual matter and expressive interaction with virtual vibrating bodies. However, placing these tools in the hands of the composer or musician has historically posed challenges in terms of (a) the computational expense of most real-time physically-based synthesis methods, (b) the difficulty of implementing these methods into modular tools that allow for the intuitive design of virtual instruments, without expert physics and/or computing knowledge, and (c) the generally limited access to such tools within popular software environments for musical creation. To this end, a set of open-source tools for designing and computing mass-interaction networks for physically-based sound synthesis is presented. The audio synthesis is performed within Max/MSP using the gen~ environment, allowing for simple model design, efficient calculation of

systems containing single-sample feedback loops, as well as extensive real-time control of physical parameters and model attributes. Through a series of benchmark examples, we exemplify various virtual instruments and interaction designs.

S5.4. Combining Texture-Derived Vibrotactile Feedback, Concatenative Synthesis, and Photogrammetry for Virtual Reality Rendering

Eduardo Magãlhaes, Emil Rosenlund Høeg, Gilberto Bernardes, Jon Ram BruunPedersen, Stefania Seran, and Rolf Nordhal

This paper describes a novel framework for real-time sonification of surface textures in virtual reality (VR), aimed towards realistically representing the experience of driving over a virtual surface. A combination of capturing techniques of real-world surfaces are used for mapping 3D geometry, texture maps, or auditory attributes (aural and vibrotactile) feedback. For the sonification rendering, we propose the use of information from primarily graphical texture features, to define target units in concatenative sound synthesis. To foster models that go beyond the current generation of simple sound textures (e.g., wind, rain, fire), towards highly "synchronized" and expressive scenarios, our contribution draws a framework for the higher-level modeling of a bicycle's kinematic rolling on ground contact, with enhanced perceptual symbiosis between auditory, visual, and vibrotactile stimuli. We scanned two surfaces represented as texture maps, consisting of different features, morphology, and matching navigation. We define target trajectories in a two-dimensional audio feature space, according to a temporal model and morphological attributes of the surfaces. This synthesis method serves two purposes: A real-time auditory feedback, and vibrotactile feedback induced through playing back the concatenated sound samples using a vibrotactile inducer speaker.

S5.5. Percussion Synthesis using Loopback Frequency Modulation Oscillators

Jennifer Hsu and Tamara Smyth

In this work, we apply recent research results in loopback frequency modulation (FM) to real-time parametric synthesis of percussion sounds. Loopback FM is a variant of FM synthesis whereby the carrier oscillator "loops back" to serve as a modulator of its own frequency. Like FM, more spectral components emerge, but further, when the loopback coefficient is made time varying, frequency trajectories that resemble the nonlinearities heard in acoustic percussion instruments appear. Here, loopback FM is used to parametrically synthesize this effect in struck percussion instruments, known to exhibit frequency sweeps (among other nonlinear characteristics) due to modal coupling. While many percussion synthesis models incorporate such nonlinear effects while aiming for acoustic accuracy, computational efficiency is often sacrificed, prohibiting real-time use. This work seeks to develop a real-time percussion synthesis model that creates a variety of novel sounds and captures the sonic qualities of nonlinear percussion instruments. A linear modal synthesis percussion model is modified to use loopback FM oscillators, which allows the model to create rich and abstract percussive hits in real-time. Musically intuitive parameters for the percussion model are emphasized, resulting in a usable percussion sound synthesizer.

S5.6. Deep Linear Autoregressive Model for Interpretable Prediction of Expressive Tempo

Akira Maezawa

Anticipating a human musician's tempo for a given piece of music using a predictable model is important for interactive music applications, but existing studies base such an anticipation on hand-crafted features. Based on recent trends in using deep learning for music performance rendering, we present an online method for multi-step prediction of the tempo curve, given the past history of tempo curves and the music score that the user is playing. We present a linear autoregressive model whose parameters are determined by a deep convolutional neural network whose input is the music score and the history of the tempo curve; such an architecture allows the machine to acquire music

performance idioms based on musical contexts, while being able to predict the timing based on the user's playing. Evaluations show that our model is capable of improving the tempo estimate over a commonly-used baseline for tempo prediction by 18%.

S5.7. Metrics for the Automatic Assessment of Music Harmony Awareness in Children

Federico Avanzini, Adriano Baratè, Luca Andrea Ludovico, and Marcella Mandanici

In the context of a general research question about the effectiveness of computer-based technologies applied to early music-harmony learning, this paper proposes a web-based tool to foster and quantitatively measure harmonic awareness in children. To this end, we have developed a web interface where young learners can listen to the leading voice of well-known music pieces and associate chords to it. During the activity, their actions can be monitored, recorded, and analyzed. An early experimentation involved 45 school teachers, whose performances have been measured in order to get user-acceptance opinions from domain experts and to determine the most suitable metrics to conduct automated performance analysis. This paper focuses on the latter aspect and proposes a set of candidate metrics to be used for future experimentation with children.

4.10. S6. Oral Session 6: Music Information Processing

Session Chair: Roger Dannenberg

S6.1. Learning to Generate Music with BachProp

Florian Colombo, Johanni Brea, and Wulfram Gerstner

As deep learning advances, algorithms of music composition increase in performance. However, most of the successful models are designed for specific musical structures. Here, we present BachProp, an algorithmic composer that can generate music scores in many styles given sufficient training data. To adapt BachProp to a broad range of musical styles, we propose a novel representation of music and train a deep network to predict the note transition probabilities of a given music corpus. In this paper, new music scores generated by BachProp are compared with the original corpora as well as with different network architectures and other related models. A set of comparative measures is used to demonstrate that BachProp captures important features of the original datasets better than other models and invites the reader to a qualitative comparison on a large collection of generated songs.

S6.2. Off-line Score Alignment for Realistic Music Practice

Yucong Jiang, Fiona Ryan, David Cartledge, and Christopher Raphael

In a common music practice scenario, a player works with a musical score, but may jump arbitrarily from one passage to another in order to drill on difficult technical challenges or pursue some other agenda requiring non-linear movement through the score. In this work, we treat the associated score alignment problem in which we seek to align a known symbolic score to the audio of the musician's practice session, identifying all "do-overs" and jumps. The result of this effort facilitates a quantitative view of a practice session, allowing feedback on coverage, tempo, tuning, rhythm, and other aspects of practice. If computationally feasible, we would prefer a globally optimal dynamic programming search strategy; however, we find such schemes only barely computationally feasible in the cases we investigate. Therefore, we develop a computationally efficient off-line algorithm suitable for practical application. We present examples, analyzing unsupervised and unscripted practice sessions on clarinet, piano, and viola, providing numerical evaluation of our score-alignment results on hand-labeled ground-truth audio data, as well as more subjective and easy-to-interpret visualizations of the results.

S6.3. Piano Score—Following by Tracking Note Evolution

Yucong Jiang and Christopher Raphael

Score following matches musical performance audio with its symbolic score in an on-line fashion. Its applications are meaningful in music practice, performance, education, and composition. This paper focuses on following piano music—one of the most challenging cases. Motivated by the time-changing features of a piano note during its lifetime, we propose a new method that models the evolution of a note in spectral space, aiming to provide an adaptive, hence better, data model. This new method is based on a switching Kalman filter in which a hidden layer of continuous variables tracks the energy of the various note harmonics. The result of this method could potentially benefit applications in de-soloing, sound synthesis, and virtual scores. This paper also proposes a straightforward evaluation method. We conducted a preliminary experiment on a small dataset of 13 min of music, consisting of 15 excerpts of real piano recordings from eight pieces. The results show the promise of this new method.

S6.4. Adaptive Score—Following System by Integrating Gaze Information

Kaede Noto, Yoshinari Takegawa, and Keiji Hirata

In actual piano practice, people of different skill levels exhibit different behaviors, for instance, leaping forward or to an upper staff, miskeying, repeating, and so on. However, many of the conventional score following systems hardly adapt to such accidental behaviors depending on individual skill level, because conventional systems usually learn the frequent or general behaviors. We develop a score-following system that can adapt a user's individuality by combining keying information with gaze, because it is well-known that the gaze is a highly reliable means of expressing a performer's thinking. Since it is difficult to collect a large amount of piano performance data reflecting individuality, we employ the framework of the Bayesian inference to adapt individuality. That is, to estimate the user's current position in piano performance, keying and gaze information are integrated into a single Bayesian inference by a Gaussian mixture model (GMM). Here, we assume both the keying and gaze information conform to normal distributions. Experimental results show that, taking into account the gaze information, our score-following system can properly cope with repetition and leaping to an upper row of a staff, in particular.

S6.5. Alternative Measures: A Musicologist Workbench for Popular Music

Beach Clark and Claire Arthur

The objective of this project is to create a digital "workbench" for quantitative analysis of popular music. The workbench is a collection of tools and data that allow for efficient and effective analysis of popular music. This project integrates software from pre-existing analytical tools, including music21, but adds methods for collecting data about popular music. The workbench includes tools that allow analysts to compare data from multiple sources. Our working prototype of the workbench contains several novel analytical tools which have the potential to generate new musicological insights through the combination of various datasets. This paper demonstrates some of the currently available tools as well as several sample analyses and features computed from this data that support trend analysis. A future release of the workbench will include a user-friendly UI for non-programmers.

4.11. P3. Poster Session 3

Session Chair: Jean Bresson

P3.1. Autoencoders for Music Sound Modeling: A Comparison of Linear, Shallow, Deep, Recurrent, and Variational Models

Fanny Roche, Thomas Hueber, Samuel Limier, and Laurent Girin

This study investigates the use of non-linear unsupervised dimensionality reduction techniques to compress a music dataset into a low-dimensional representation which can be used in turn

for the synthesis of new sounds. We systematically compare (shallow) autoencoders (AEs), deep autoencoders (DAEs), recurrent autoencoders (with long short-term memory cells—LSTM-AEs) and variational autoencoders (VAEs) with principal component analysis (PCA) for representing the high-resolution short-term magnitude spectrum of a large and dense dataset of music notes into a lower-dimensional vector (and then convert it back to a magnitude spectrum used for sound resynthesis). Our experiments were conducted on the publicly available multi-instrument and multi-pitch database, NSynth. Interestingly, and contrary to the recent literature on image processing, we can show that PCA systematically outperforms shallow AE. Only deep and recurrent architectures (DAEs and LSTMAEs) lead to a lower reconstruction error. The optimization criterion in VAEs being the sum of the reconstruction error and a regularization term, it naturally leads to a lower reconstruction accuracy than DAEs but we show that VAEs are still able to outperform PCA while providing a low-dimensional latent space with nice "usability" properties. We also provide corresponding objective measures of perceptual audio quality (PEMO-Q scores), which generally correlate well with the reconstruction error.

P3.2. Polytopic Reconfiguration: A Graph-Based Scheme for the Multiscale Transformation of Music Segments and its Perceptual Assessment

Valentin Gillot and Frédéric Bimbot

Music is a sequential process for which relations between adjacent elements play an important role. Expectation processes based on alternations of similarity and novelty contribute to the structure of the musical flow. In this work, we explore a polytopic representation of music, which accounts for expectation systems developing at several timescales in parallel. After recalling properties of polytopic representations for describing multi-scale implication processes, we introduce a scheme for recomposing musical sequences by simple transformations of their support polytope. A specific set of permutations (referred to as primer preserving permutations or PPP) are of particular interest, as they preserve systems of analogical implications within musical segments. By means of a perceptual test, we study the impact of PPP-based transformations by applying them to the choruses of pop songs in midi format and comparing the result with randomly generated permutations (RGP). In our test, subjects are asked to rate musical excerpts reconfigured by PPP-based transformations versus RGP-based ones in terms of musical consistency and of attractiveness. Results indicate that PPP-transformed segments score distinctly better than RGP-transformed for the two criteria, suggesting that the preservation of implication systems plays an important role in the subjective acceptability of the transformation. Additionally, from the perspective of building an automatic recomposition system for artistic creation purposes, we introduce, in the appendix, the preliminary version of an automatic method for decomposing segments into low-scale musical elements, taking into account possible phase-shifts between the musical surface of the melody and the metrical information.

P3.3. Non-Linear Contact Sound Synthesis for Real-Time Audio-Visual Applications using Modal Textures

Martin Maunsbach and Stefania Seran

Sound design is an integral part of making a virtual environment come to life. Spatialization is important to the perceptual localization of sounds, while the quality determines how well virtual objects come to life. The implementation of pre-recorded audio for physical interactions in virtual environments often requires a vast library of audio files to distinguish each interaction from the other. This paper explains the implementation of a modal synthesis toolkit for the Unity game engine to automatically add impact and rolling sounds to interacting objects. Position-dependent sounds are achieved using a custom shader that can contain textures with modal weighting parameters. The two types of contact sounds are synthesized using a mechanical oscillator describing a mass-spring system. We describe the discretization methods adopted, the solution of the nonlinear interaction, and an implementation in the Unity game engine.

P3.4. Analysis of Vocal Ornamentation in Iranian Classical Music

Sepideh Shaei

In this paper, we study tahrir, a melismatic vocal ornamentation which is an essential characteristic of Persian classical music and can be compared to yodeling. It is considered the most important technique through which the vocalist can display his/her prowess. In Persian, the nightingale's song is used as a metaphor for tahrir and sometimes for a specific type of tahrir. Here, we examine tahrir through a case study. We have chosen two prominent singers of Persian classical music, one contemporary and one from the 20th century. In our analysis, we have appropriated both audio recordings and transcriptions by one of the most prominent ethnomusicologists, Masudiyeh, who has worked on the music of Iran. This paper is the first step towards the computational modeling and recognition of different types of tahrirs. Here, we have studied two types of tahrirs, mainly nashib and farāz, and their combination through three different performance samples by two prominent vocalists. More than 20 types of tahrirs have been identified by Iranian musicians and music theorists. We are currently working on developing a method to computationally identify these models.

P3.5. VUSAA: An Augmented Reality Mobile App for Urban Soundwalks

Josué Moreno and Vesa Norilo

This paper presents VUSAA, an augmented reality soundwalking application for Apple iOS devices. The application is based on the idea of urban sonic acupuncture, providing site-aware generative audio content aligned with the present sonic environment. The sound-generating algorithm was implemented in Kronos, a declarative programming language for musical signal processing. We discuss the conceptual framework and implementation of the application, along with the practical considerations of deploying it via a commercial platform. We present results from a number of soundwalks so far organized and outline an approach to develop new models for urban dwelling.

P3.6. A Framework for Multi-f0 Modeling in SATB Choirs

Helena Cuesta, Emilia Gómez, and Pritish Chandna

Fundamental frequency (f0) modeling is an important but relatively unexplored aspect of choir singing. Performance evaluation as well as auditory analysis of singing, whether individually or in a choir, often depend on extracting f0 contours for the singing voice. However, due to the large number of singers singing at a similar frequency range, extracting the exact individual pitch contours from choir recordings is a challenging task. In this paper, we address this task and develop a methodology for modeling pitch contours of SATB choir recordings. A typical SATB choir consists of four parts, each covering a distinct range of pitches and often with multiple singers each. We first evaluate some state-of-the-art multi-f0 estimation systems for the particular case of choirs with a single singer per part, and observe that the pitch of individual singers can be estimated to a relatively high degree of accuracy. We observe, however, that the scenario of multiple singers for each choir part (i.e., unison singing) is far more challenging. In this work, we propose a methodology based on combining a multi-f0 estimation methodology based on deep learning followed by a set of traditional DSP techniques to model f0 and its dispersion instead of a single f0 trajectory for each choir part. We present and discuss our observations and test our framework with different singer configurations.

P3.7. Representations of Self-Coupled Modal Oscillators with Time-Varying Frequency

Tamara Smyth and Jennifer Hsu

In this work, we examine a simple mass-spring system in which the natural frequency is modulated by its own oscillations, a self-coupling that creates a feedback system in which the output signal "loops back" with an applied coefficient to modulate the frequency. This system is first represented as a mass-spring system, then as an extension of well-known frequency modulation synthesis (FM) coined

"loopback FM", and finally, as a closed-form representation that has a form similar to the transfer function of a "stretched" allpass filter with a time-varying delay, but with the fundamental difference that it is used here as a time-domain signal, the real part of which is the sounding waveform. This final representation allows for integration of the instantaneous frequency in the FM representation and ultimately a mapping from its parameters to those of loopback FM. In addition to predicting the sounding frequency (pitch glides) of loopback FM for a given carrier frequency and time-varying loopback coefficient, or equivalently of the self-coupled oscillator for a given natural frequency and coupling coefficient, the closed form representation is seen to be a more accurate representation of the system as it does not introduce a unit-sample delay in the feedback loop, nor is it as numerically sensitive to the sampling rate.

P3.8. SonaGraph. A Cartoonified Spectral Model for Music Composition

Andrea Valle

This paper presents SonaGraph, a framework and an application for a simplified but efficient harmonic spectrum analyzer suitable for assisted and algorithmic composition. The model is inspired by the analog, Sonagraph, and relies on a constant-Q bandpass filter bank. First, the historical Sonagraph is introduced. Then, starting from it, a simplified ("cartoonified") model is discussed. An implementation in SuperCollider is presented that includes various utilities (interactive GUIs, music notation generation, graphic export, data communication). A comparison of results in relation to other tools for assisted composition is presented. Finally, some musical examples are discussed that make use of spectral data from SonaGraph to generate, retrieve, and display music information.

P3.9. Sound in Multiples: Synchrony and Interaction Design of Coupled Oscillator Networks

Nolan Lem

Systems of coupled oscillators can be employed in a variety of algorithmic settings to explore the self-organizing dynamics of synchronization. In the realm of audio-visual generation, coupled oscillator networks can be usefully applied to musical content related to sound synthesis, rhythmic generation, and compositional design. By formulating different models of these generative dynamical systems, I outline different methodologies from which to generate sound from collections of interacting oscillators and discuss how their rich non-linear dynamics can be exploited in the context of sound-based art. A summary of these mathematical models is discussed and a range of applications are proposed in which they may be useful in producing and analyzing sound. I discuss these models in relationship to one of my own kinetic sound sculptures to analyze to what extent they can be used to characterize synchrony as an analytical tool.

P3.10. Jazz Mapping an Analytical and Computational Approach to Jazz Improvisation

Dimitrios Vassilakis, Anastasia Georgaki, and Christina Anagnostopoulou

"Jazz mapping" is a multi-layered analytical approach to jazz improvisation. It is based on hierarchical segmentation and the categorization of segments, or constituents, according to their function in the overall improvisation. The approach aims at identifying higher-level semantics of transcribed and recorded jazz solos. At these initial stages, analytical decisions are rather exploratory and rely on the input of one of the authors and an experienced jazz performer. We apply the method to two well-known solos, by Sonny Rollins and Charlie Parker, and discuss how improvisations resemble story-telling, employing a broad range of structural, expressive, and technical tools, usually associated with linguistic production, experience, and meaning. We elucidate the implicit choices of experienced jazz improvisers, who have developed a strong command over the language and can communicate expressive intent, elicit emotional responses, and unfold musical "stories" that are memorable and enjoyable to fellow musicians and listeners. We also comment on potential artificial intelligence applications of this work to music research and performance.

P3.11. Visual Pitch Estimation

Sophia Koepke, Olivia Wiles, and Andrew Zisserman

In this work, we propose the task of automatically estimating pitch (fundamental frequency) from video frames of violin playing using vision alone. Here, we consider only monophonic violin playing (where only one note is being played at a time). In order to investigate this task, we curate a new dataset of monophonic violin playing. We propose a convolutional neural network (CNN) architecture that is trained using a student–teacher strategy to distil knowledge from the audio domain to the visual domain. At test time, our network takes video frames as input and directly regresses the pitch. We train and test this architecture on different subsets of our new dataset. We show that this task (i.e., pitch prediction from vision) is actually possible. Furthermore, we verify that the network has indeed learnt to focus on salient parts of the image, e.g., the left hand of the violin player is used as a visual cue to estimate pitch.

4.12. D3. Demo Session 3

Session Chair: Ana M. Barbancho

D3.1. Miningsuite: A Comprehensive Matlab Framework for Signal, Audio, and Music Analysis, Articulating Audio and Symbolic Approaches

Olivier Lartillot

The MiningSuite is a free open-source and comprehensive Matlab framework for the analysis of signals, audio recordings, music recordings, music scores, and other signals, such as motion capture data, etc., under a common modular framework. It adds a syntactic layer on top of Matlab, so that advanced operations can be specified using a simple and adaptive syntax. This makes the Matlab environment very easy to use for beginners, and at the same time allows power users to design complex workflows in a modular and concise way through a simple assemblage of operators featuring a large set of options. The MiningSuite is an extension of MIRtoolbox, a Matlab toolbox that has become a reference tool in MIR.

D3.2. Drawing Geometric Figures with Braille Description through a Speech Recognition System

África Chamorro, Ana M. Barbancho, Isabel Barbancho, and Lorenzo J. Tardón

In this contribution, a system that represents drawings of geometric figures along with their description transcribed in Braille controlled by means of commands acquired by a speech recognition scheme is presented. The designed system recognizes the spoken descriptions needed to draw simple geometric objects: Shape, color, size, and position of the figures in the drawing. The speech recognition method selected is based on a distance measure defined with Mel-frequency cepstral coefficients (MFCCs). The complete system can be used by both people with visual and hearing impairments thanks to its interface which, in addition to showing the drawing and the corresponding transcription in Braille, also allows the user to hear the description of commands and the final drawing.

D3.3. Interactive Music Training System

Daniel Moreno, Isabel Barbancho, Ana M. Barbancho, and Lorenzo J. Tardón

In this contribution, we present an interactive system for playing while learning music. The game is based on different computer games controlled by the user with a remote control. The remote control has been implemented using inertial measurement sensors (IMU) for 3D tracking. The computer games are programmed in Python and allow the practice of rhythm as well as the tune, ascending or descending, of musical notes.

D3.4. Copying Clave—A Turing Test

Simon Blackmore

A blindfolded instructor (evaluator) plays a clave pattern. A computer captures and repeats the pattern. After 1 min, the experiment stops. This process is repeated by a human who also tries to copy the clave. After another minute, they stop and the evaluator assesses both performances.

D3.5. Resonance Improviser: A System for Transmitting the Embodied Sensations of Vocalization Between Two People During Improvisation

Tejaswinee Kelkar and Lynda Joy Gerry

This is a system prototype for joint vocal improvisation between two people that involves the sharing of embodied sensations of vocal production. This is accomplished by using actuators that excite two participants' rib cages with each other's voices, turning a person's body into a loud speaker. A microphone transmits vocal signals and the players are given a Max Patch to modulate the sound and feel of their voice. The receiver hears the other person's speech and effects through their own body (as if it were their own voice), while also feeling the resonance of the sound signal as it would resonate in the chest cavity of the other. The two players try to re-enact and improvise a script prompt provided to them while not knowing what the other person can hear of their voice. The game may or may not turn collaborative, adversarial, or artistic depending on the game play.

D3.6. Finding New Practice Material through Chord-Based Exploration of a Large Music Catalogue

Johan Pauwels and Mark B. Sandler

Our demo is a web app that suggests new practice material to music learners based on automatic chord analysis. It is aimed at music practitioners of any skill set, playing any instrument, as long as they know how to play along with a chord sheet. Users need to select a number of chords in the app and are then presented with a list of music pieces containing those chords. Each of those pieces can be played back while its chord transcription is displayed in sync to the music. This enables a variety of practice scenarios, ranging from following the chords in a piece to using the suggested music as a backing track to practice soloing over.

D3.7. Internal Complexity for Exploratory Interaction

Mads Hobye

When designing interactive sound for non-utilitarian ludic interaction, internal complexity can be a way of opening up a space for curiosity and exploration. Internal complexity should be understood as non-linear mappings between the input and the parameters they affect in the output (sound). This paper presents three different experiments which explore ways to create internal complexity with simple interfaces for curious exploration.

D3.8. Adaptive Body Movement Sonication in Music and Therapy

Christian Baumann, Johanna Friederike Baarlink, and Jan-Torsten Milde

In this paper, we describe the ongoing research on the development of a body movement sonification system. High precision and high resolution wireless sensors are used to track the body movement and record muscle excitation. We are currently using six sensors. In the final version of the system, full body tracking can be achieved. The recording system provides a web server, including a simple REST API, which streams the recorded data in JSON format. An intermediate proxy server preprocesses the data and transmits it to the final sonification system. The sonification system is implemented using the web audio api. We are experimenting with a set of different sonification strategies and algorithms. Currently, we are testing the system as part of an interactive guided therapy, establishing additional acoustic feedback channels for the patient. In a second stage of the research,

we are going to use the system in a more musical and artistic way. More specifically, we plan to use the system in cooperation with a violist, where the acoustic feedback channel will be integrated into the performance.

4.13. S7. Oral Session 7: Multimodality and (e)Motions

Session Chair: Sofia Dahl

S7.1. VocalistMirror: A Singer Support Interface for Avoiding Undesirable Facial Expressions

Kin Wah Edward Lin, Tomoyasu Nakano, and Masataka Goto

We present VocalistMirror, an interactive user interface that enables a singer to avoid their undesirable facial expressions in singing video recordings. Since singers usually focus on singing expressions and do not care about facial expressions, when watching singing videos they recorded, they sometimes notice that some of their facial expressions are undesirable. VocalistMirror allows a singer to first specify their undesirable facial expressions in a recorded video, and then sing again while seeing a real-time warning that is shown when the facial expression of the singer becomes similar to one of the specified undesirable expressions. It also displays karaoke-style lyrics with a piano-roll melody and visualizes the acoustic features of singing voices. The iOS ARKit framework is used to quantify the facial expression as a 52-dimensional vector, which is then used to compute the distance from undesirable expressions. Our experimental results showed the potential of the proposed interface.

S7.2. Audiovisual Perception of Arousal, Valence, and Effort in Contemporary Cello Performance

Hanna Järveläinen

Perceived arousal, valence, and effort were measured continuously from auditory, visual, and audiovisual cues using a recorded performance of a contemporary cello piece. Effort (perceived exertion of the performer) was added for two motivations: To investigate its potential as a measure and its association with arousal in audiovisual perception. Fifty-two subjects participated in the experiment. Results were analyzed using activity analysis and functional data analysis. Arousal and effort were perceived with significant coordination between participants from auditory, visual, as well as audiovisual cues. Significant differences were detected between auditory and visual channels but not between arousal and effort. Valence, in contrast, showed no significant coordination between participants. The relative importance of the visual channel is discussed.

S7.3. Dancing Dots—Investigating the Link between the Dancer and Musician in Swedish Folk Dance

Olof Misgeld, Andre Holzapfel, and Sven Ahlbäck

The link between musicians and dancers is generally described as strong in much traditional music and this holds also for Scandinavian folk music—spelmansmusik. Understanding the interaction of music and dance has potential for the development of theories of performance strategies in artistic practice and for the development of interactive systems. In this paper, we investigate this link by having Swedish folk musicians perform to animations generated from motion capture recordings of dancers. The different stimuli focus on the motions of selected body parts as moving white dots on a computer screen with the aim to understand how different movements can provide reliable cues for musicians. Sound recordings of fiddlers playing to the "dancing dot" were analyzed using automatic alignment to the original music performance related to the dance recordings. Interviews were conducted with musicians and comments were collected in order to shed light on strategies when playing for dancing. Results illustrate a reliable alignment to renderings showing full skeletons of dancers, and an advantage of focused displays of movements in the upper back of the dancer.

4.14. S8. Oral Session 8: Machine Learning

Session Chair: Olivier Lartillot

S8.1. Conditioning a Recurrent Neural Network to Synthesize Musical Instrument Transients

Lonce Wyse and Muhammad Huzaifah

A recurrent neural network (RNN) is trained to predict sound samples based on audio input augmented by control parameter information for pitch, volume, and instrument identification. During the generative phase following training, audio input is taken from the output of the previous time step, and the parameters are externally controlled, allowing the network to be played as a musical instrument. Building on an architecture developed in previous work, we focus on the learning and synthesis of transients—the temporal response of the network during the short time (tens of milliseconds) following the onset and offset of a control signal. We find that the network learns the particular transient characteristics of two different synthetic instruments, and furthermore shows some ability to interpolate between the characteristics of the instruments used in training in response to novel parameter settings. We also study the behavior of the units in hidden layers of the RNN using various visualization techniques and find a variety of volume-specific response characteristics.

S8.2. Predicting Perceived Dissonance of Piano Chords Using a ChordClass Invariant CNN and Deep Layered Learning

Juliette Dubois, Anders Elowsson, and Anders Friberg

This paper presents a convolutional neural network (CNN) able to predict the perceived dissonance of piano chords. Ratings of dissonance for short audio excerpts were combined from two different datasets and groups of listeners. The CNN uses two branches in a directed acyclic graph (DAG). The first branch receives input from a pitch estimation algorithm, restructured into a pitch chroma. The second branch analyses interactions between close partials, known to affect our perception of dissonance and roughness. The analysis is pitch invariant in both branches, facilitated by convolution across log-frequency and octave wide max-pooling. Ensemble learning was used to improve the accuracy of the predictions. The coefficient of determination ($R2$) between ratings and predictions is close to 0.7 in a cross-validation test of the combined dataset. The system significantly outperforms recent computational models. An ablation study tested the impact of the pitch chroma and partial analysis branches separately, concluding that the deep layered learning approach with a pitch chroma was driving the high performance.

S8.3. Belief Propagation Algorithm for Automatic Chord Estimation

Vincent P. Martin, Sylvain Reynal, Dogac Basaran, and Hélène-Camille Crayencour

This work aims at bridging the gap between two completely distinct research fields: Digital communications and music information retrieval (MIR). While works in the MIR community have long used algorithms borrowed from speech signal processing, text recognition, or image processing, to our knowledge, very scarce works based on digital communications algorithms have been produced. This paper specifically targets the use of the belief propagation algorithm for the task of automatic chord estimation. This algorithm has widespread use in iterative decoders for error correcting codes and we show that it offers improved performances in ACE by genuinely incorporating the ability to take constraints between distant parts of the song into account. It certainly represents a promising alternative to traditional MIR graphical models approaches, in particular hidden Markov models.

S8.4. HMM-Based Glissando Detection for Recordings of the Chinese Bamboo Flute

Changhong Wang, Emmanouil Benetos, Xiaojie Meng, and Elaine Chew

Playing techniques, such as ornamentations and articulation effects, constitute important aspects of music performance. However, their computational analysis is still at an early stage due to a lack

of instrument diversity, established methodologies, and informative data. Focusing on the Chinese bamboo flute, we introduce a two-stage glissando detection system based on hidden Markov models (HMMs) with Gaussian mixtures. A rule-based segmentation process extracts glissando candidates that are consecutive note changes in the same direction. Glissandi are then identified by two HMMs. The study uses a newly created dataset of Chinese bamboo flute recordings, including both isolated glissandi and real-world pieces. The results, based on both frame- and segment-based evaluation for ascending and descending glissandi, respectively, confirm the feasibility of the proposed method for glissando detection. Better detection performance of the ascending glissandi over descending ones is obtained due to their more regular patterns. Inaccurate pitch estimation forms a main obstacle for successful fully-automated glissando detection. The dataset and method can be used for performance analysis.

S8.5. Towards CNN-Based Acoustic Modeling of Seventh Chords for Automatic Chord Recognition

Christon-Ragavan Nadar, Jakob Abeßer, and Sascha Grollmisch

In this paper, we build upon a recently proposed deep convolutional neural network architecture for automatic chord recognition (ACR). We focus on extending the commonly used major/minor vocabulary (24 classes) to an extended chord vocabulary of seven chord types with a total of 84 classes. In our experiments, we compare joint and separate classification of the chord type and chord root pitch class using one or two separate models, respectively. We perform a large-scale evaluation using various combinations of training and test sets of different timbre complexity. Our results show that ACR with an extended chord vocabulary achieves high f-scores of 0.97 for isolated chord recordings and 0.66 for mixed contemporary popular music recordings. While the joint ACR modeling leads to the best results for isolated instrument recordings, the separate modeling strategy performs best for complex music recordings. Alongside this paper, we publish a novel dataset for extended-vocabulary chord recognition which consists of synthetically generated isolated recordings of various musical instruments.

S8.6. From Jigs and Reels to Schottisar och Polskor: Generating Scandinavian-Like Folk Music with Deep Recurrent Networks

Eric Hallström, Simon Mossmyr, Bob L. Sturm, Victor Hansjons Vegeborn, and Jonas Wedin

The use of recurrent neural networks for modeling and generating music has been shown to be quite effective for compact textual transcriptions of traditional music from Ireland and the UK. We explore how well these models perform for textual transcriptions of traditional music from Scandinavia. This type of music has characteristics that are similar to and different from that of Irish music, e.g., mode, rhythm, and structure. We investigate the effects of different architectures and training regimens, and evaluate the resulting models using three methods: A comparison of statistics between real and generated transcriptions, an appraisal of generated transcriptions via a semi-structured interview with an expert in Swedish folk music, and an exercise conducted with students of Scandinavian folk music. We find that some of our models can generate new transcriptions sharing characteristics with Scandinavian folk music, but which often lack the simplicity of real transcriptions. One of our models has been implemented online at http://www.folkrnn.org for anyone to try.

S8.7. Modeling and Learning Rhythm Structure

Francesco Foscarin, Florent Jacquemard, and Philippe Rigaux

We present a model to express preferences on rhythmic structure, based on probabilistic context-free grammars, and a procedure that learns the grammars' probabilities from a dataset of scores or quantized MIDI files. The model formally defines rules related to rhythmic subdivisions and durations that are in general given in an informal language. Rule preference is then specified with probability values. One targeted application is the aggregation of rule probabilities to qualify an entire rhythm for tasks, like automatic music generation and music transcription. The paper also reports an application of this approach on two datasets.

Conflicts of Interest: The authors declare no conflict of interest.