DMRN+14: Digital Music Research Network One-day Workshop 2019

Digital Music
Research Network

Arts Two Lecture Theatre

Queen Mary University of London

Tuesday 17th December 2019

Chairs

Panos Kudumakis and Simon Dixon

Queen Mary
University of London

centre for digital music

Programme

10:30	**Registration** opens **Tea/Coffee**
11:00	**Welcome** and opening remarks **Prof. Simon Dixon** (Centre for Digital Music, Queen Mary University of London)
11:10	**KEYNOTE** “Data and Human Interpretation: Music and Beyond", **Dr Cynthia Liem (Delft University of Technology)**
11:50	“Software for Analysis of Harmony as a Computer Tool for Search of Tonal Cadencies in Various Midi Files”, **Ferková Eva and Šukola Michal (Academy of Performing Arts, Bratislava, Slovakia)**
12:10	**“**Human vs. Automated Judgements of Similarity in a Global Music Sample”, **Hideo Daikoku and Ding Shenghao (Keio University, Japan), Ujwal Sriharsha Sanne (Queen Mary University of London), Marino Kinoshita, Rei Konno, Yoichi Kitayama, Shinya Fujii and Patrick E. Savage (Keio University, Japan)**
12:30	“Modelling Keys and Modulation with Scales and Harmonic Progressions”, **Laurent Feisthauer, Louis Bigo, Mathieu Giraud (Université de Lille, France) and Florence Levé Université de Picardie Jules Verne & Université de Lille, France)**
12:50	**Tea/Coffee** **Posters** will be on display
14:00	“Exploring the Aesthetics and Utility of Sonification with Isomorph– Interactive Sonification for Molecular Physics”, **Joseph Hyde (Bath Spa University), Thomas J. Mitchell (University of the West of England), Helen M. Deeks, David R. Glowacki and Alex J. Jones (University of Bristol)**
14:20	“Alter: An Ensemble Work Composed with and about AI”, **David De Roure (University of Oxford & The Alan Turing Institute), Emily Howard, Robert Laidlow (Royal Northern College of Music) and Pip Willcox (University of Oxford & The National Archives)**
14:40	“Unmixer: An Interface for Extracting and Remixing Loops”, **Jordan B. L. Smith, Yuta Kawasaki and Masataka Goto (National Institute of Advanced Industrial Science and Technology, Japan)**
15:00	**“**The Impact of Dataset Modifications on Music Similarity Measures”, **Roberto Piassi Passos Bodo (University of São Paulo), Emmanouil Benetos (Queen Mary University of London) and Marcelo Queiroz (University of São Paulo)**
15:20	**Tea/Coffee** **Posters** will be on display
15:40	**“**Dig That Lick: Exploring Patterns in Jazz Solos”, **Simon Dixon, Polina Proutskova (Queen Mary University of London, UK), Tillman Weyde, Daniel Wolff (City University of London, UK), Martin Pfleiderer, Klaus Frieler, Frank Höger (University of Music Weimar, Germany), Hélène-Camille Crayencour, Jordan B. L. Smith (National Center for Scientific Research, France), Geoffroy Peeters, Doğaç Başaran (IRCAM, France), Gabriel Solis, Lucas Henry (University of Illinois, USA), Krin Gabbard, Andrew Vogel (Columbia University, USA)**
16:00	“Modulation Spectra for Musical Dynamics Perception and Retrieval”, **Luca Marinelli, Athanasios Lykartsis (Technischen Universität Berlin) and Charalampos Saitis (Queen Mary University of London)**
16:20	“Searching for Efficient Processing Pipelines Applied to MIR in Embedded Systems”, **Filipe Lins, Marcelo Johann and Rodrigo Schramm (UFRGS, Brazil)**
16:40	**“**Homepage and Search Personalization at Spotify”, **Mi Tian, Rishabh Mehrotra, Lucas Maystre and Mounia Lalmas (Spotify, UK)**
17:00	**Panel Discussion**
17:30	**Close***

* - There will be an opportunity to continue discussions after the Workshop in a nearby Pub/Restaurant.

Posters

1	**“**Automatic Music Accompaniment with a Chroma-based Music Data Representation”, **Lele Liu and Emmanouil Benetos (Queen Mary University of London)**
2	“A MELD TimeMachine for Wagner’s Lohengrin”, **David Lewis, Kevin Page and Laurence Dreyfus (University of Oxford)**
3	“RadioMe: Artificially Intelligent Radio for People with Dementia”, **Satvik Venkatesh, David Moffat, and Eduardo Reck Miranda (University of Plymouth)**
4	**“**A Decentralized MELD Agent Framework for Computational Analysis of the Live Music Archive”, **Graham Klyne (University of Oxford), Thomas Wilmering (Queen Mary University of London), John Pybus and Kevin Page (University of Oxford)**
5	**“**AMT for Musicians: Performed-MIDI-to-Score Transcription”, **Francesco Foscarin (CNAM, Paris), Florent Jaquemard (CNAM and INRIA, Paris), Philippe Rigaux and Raphaël Fournier-S'niehotta (CNAM, Paris)**
6	“Using Different Feature Selection Methods for Mood Prediction”, **Cornelia Metzig and Mark Sandler (Queen Mary University of London)**
7	**“**The Algorithmically Enhanced Stylophone”, **David De Roure (University of Oxford & The Alan Turing Institute), Alan Chamberlain (University of Nottingham), Iain Emsley (University of Sussex) and John Pybus (University of Oxford)**
8	**“**Context-Aware Audio QoE: A Case Study on the Apollo 11 Audio Archive”, **Alessandro Ragano (University College Dublin, Ireland & Queen Mary University of London & The Alan Turing Institute), Emmanouil Benetos (Queen Mary University of London & The Alan Turing Institute) and Andrew Hines (Queen Mary University of London)**
9	“Retro in Digital: Understanding the Semantics of Audio Effects”, **Gary Bromham (Queen Mary University of London), David Moffat (University of Plymouth), Mathieu Barthet and György Fazekas (Queen Mary University of London)**
10	“Computational Comparison Between Different Styles of Singing Voice in Terms of the Pitch”, **Yukun Li and Simon Dixon (Queen Mary University of London)**

A Software for Analysis of Harmony as a Computer Tool for Search of Tonal Cadencies in Various Midi Files

Eva Ferková[1] and Michal Šukola[1]

[1]Department of Music Theory, Faculty of Music and Dance, Academy of Performing Arts, Slovakia
ferkova@vsmu.sk, michal.sukola@gmail.com

***Abstract*— We present the original software (as Sibelis plugin) for analysis of tonal harmony in three layers – chords, tonal keys, harmonic functions. The main features of tonal harmony in music are based on the processes, which lead to the tonal center through the progression of chords, having their harmonic/tonal functions. The highest frequented progressions are those of cadential chord-orders, which centralize the tonic triad as the most important chord of the tonal-key center. To find every harmonic-tonal cadence:**

1. **we recognize in the composition the *chord structures* (named chord classes similarly to pitch classes) and *the root of every chord***
2. **we determine the *tonal key* of every section of the music, that means we identify the first degree of the key scale and the genus of it (major, minor).**
3. **and finally we assign every chord it´s *harmonic function*.**

I. CHORD CLASS

Chord classes are defined in presented software according to textbooks for learning classic harmony [1] as triads (major, minor, diminished, augmented) and 7th chords (dominant, major, minor, diminished, half-diminished, augmented-major, minor-major). Their structure and signs for describing in scores are presented in [2], in Sibelius output in three lines.

II. TONAL KEY AND HARMONIC FUNCTION

The key is possible to find according to the key-signature of the score at the beginning. However there are many midi files on the web, which have no information about this signature, there are no accidentals used with every note separately, there are only midi numbers. The determination of tonal key in such midi file is more difficult. Therefore our software includes tools as procedures for its determination through quantity (rhythmical values, repetitions, accents etc. as "chordal weight") and quality (occurrences of major or minor triads as potential tonic, or of dominant seventh as a chord on the 5th degree, or of the half-diminished seventh as a chord on the 2nd degree, or of the diminished seventh as a chord on the 7th degree). The harmonic function is determined according to the degree-position of the root of the chord in the scale of the identified tonal key (see table I). Details of the software might be explained in detail.

TABLE I.

Harmonic functions		
The degree of the scale	***Possible chord-classes on this degree***	***Possible harmonic function***
I	Major or minor triad	T
II	Minor or diminished triad, half diminished 7th chord	either "s" or "d"as collateral functions
III	Minor or major triad or minor or major 7th chord	either "d" or "t" as collateral functions
IV	Minor or major triad or minor (major) 7th chord	S
V	Major triad or D7	D
VI	Major or minor triad or major or minor 7th chord	either "t" or "s" as collateral functions
VII	Diminished triad or diminished 7th chord, or half- diminished 7th chord	either "d" or "s" as collateral functions

III. CADENCE PROGRESSION IN DETERMINED SPACE OF TONAL KEY

The progression of harmonic functions from tonic progressing to subdominant, after that to dominant chord(s) and back to tonic is considered to be cadential,.

IV. CONCLUSION

The style of the set of compositions might be defined according to the typical or most frequented cadential progressions as a manuscript of the composer [3] or of the ethnic society (in folk music) or of the pop-music group or creators.

REFERENCES

[1] Piston, W**.** (1987): *Harmony*. Revised by DeVoto. M. New York -London: W. W. Norton & Company

[2] Ferková, E.-Šidlík, P.-Ždímal, M. (2007): Chordal Evaluation in MIDI-Based Harmonic Analysis: Mozart, Schubert, and Brahms in: *Computing in musicology 15, Tonal Theory for the Digital Age.* Stanford, p. 186-200

[3] Tymoczko, D. (2010): *What Makes Music Sound Good?* MUSIC 105, https://dmitri.mycpanel.princeton.edu/whatmakesmusicsoundgood.html

[4] Ferková, E.-Šukola, M. (2017): *Demonstration of chord detection and analysis software.* Malaga http://fma2017.uma.es/docs/FMA2017_Proceedings.pdf

Human vs. Automated Judgements of Similarity in a Global Music Sample

Hideo Daikoku[1], Ding Shenghao[1], Ujwal Sriharsha Sanne[2], Marino Kinoshita[1], Rei Konno[1], Yoichi Kitayama[1], Shinya Fujii[1], Patrick E. Savage[1]

[1*]Keio University, Shonan Fujisawa Campus, Japan, hideo-daikoku@keio.jp
[2]Queen Mary University of London, UK

Abstract— **While recent developments in MIR have proven automatic analysis of content-based similarity to be reliable, it remains unclear whether these algorithms can be meaningfully applied to cross-cultural analyses of more diverse samples. There is thus a need to evaluate automatic methods with perceptual ground truth data. Here we conducted multiple perceptual ratings tests on a subset of the *Cantometrics* recordings across 62 participants. We compared evaluating perceptual similarity by pairwise comparison, an odd-one-out method and a simplified feature evaluation based on 6 *Cantometric* features (namely, Ornamentation, Vocal Range, Tempo, Rhythmic Regularity, Vocal Tension, and Vocal Texture). Preliminary analysis shows that while the same person may disagree with themselves depending on the evaluation method, groups in general tend to agree with each other for the same method. We then compare the perceptual similarity ratings against existing automated algorithms and find minimal correlation, suggesting that there is still much room for improvement in automated cross-cultural content-based similarity.**

ACKNOWLEDGMENT

We thank Ujwal Sriharsha Sanne for extensive help implementing the automated similarity analyses. Ujwal tragically passed away only a few weeks before submission, so he was unable to read and approve the submission, but we have included him as an author in order to acknowledge his work on the project and respect his memory. We thank the YAMAHA corporation (particularly Taiki Shimozono, Rento Tanase, Satoshi Usa, and Takuya Fujishima) for providing support, feedback, and experiment participants. Additional funding for this study was provided by Startup Grants from the Keio Global Research Institute and Keio Research Institute at SFC awarded to P.E.S.

REFERENCES

[1] Michael Casey, Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney. Content-based music information retrieval: Current directions and future challenges. Proceedings of the IEEE, 96:668 – 696, 05 2008.

[2] Alan Lomax. Folk song style and culture. American Association for the Advancement of Science, 1968.

[3] Alan Lomax. Cantometrics: An Approach to the Anthropology of Music. Berkeley: Uneversity of California Extension Media Center, 1976

[4] Samuel A Mehr, Manvir Singh, Dean Knox, Daniel Ketter, Daniel Pickens-Jones, S. Atwood, Christopher Lucas, Nori Jacoby, Alena Egner, Erin J Hopkins, and et al. Universality and diversity in human song, Nov 2018.

[5] Maria Panteli, Emmanouil Benetos, and Simon Dixon. A computational study on outliers in world music. PLOS ONE, 12(12):e0189399, December 2017.

[6] Patrick E. Savage. Alan lomaxs cantometrics project: A comprehensive review. Music & Science, 1:2059204318786084, 2018.

[7] Patrick E. Savage, Steven Brown, Emi Sakai, and Thomas E. Currie. Statistical universals reveal the structures and functions of human music. Proceedings of the National Academy of Sciences, 112(29):8987–8992, 2015.

[8] Patrick E. Savage, Emily S Merritt, Tom I. Rzeszutek, and Steven O. Brown. Cantocore: A new cross-cultural song classification scheme. 2012.

[9] George Tzanetakis, W. Schloss, and Mathew Wright. Computational ethnomusicology. Journal of Interdisciplinary Music Studies, 1:1–24, 01 2007

*Research supported by Yamaha.

Modeling Keys and Modulation with Scales and Harmonic Progressions

Laurent Feisthauer[1], Louis Bigo[1], Mathieu Giraud[1] and Florence Levé[1,2]

[1]CRIStAL, UMR 9189 CNRS, Université de Lille, France, laurent.feisthauer@univ-lille.fr
[2]MIS, Université de Picardie Jules Verne, France

Abstract— **Key changes, also called modulations, are common in tonal music. With cadences, they contribute to what can be called a tonal path, and generally to the structure of music. To better model key modulations throughout a piece, we propose to take into account scales, strong harmonic progressions as well as key relationships. We test these proposals on a corpus of Mozart String Quartets movements.**

I. INTRODUCTION

Tonal music scores are built around one or more keys and their associated scales. Studying the *tonal path* of a score consists in determining keys but also the very moment when we switch from one key to another, which notes are important to the modulation and ambiguous zones.

Key finding is a well-studied topic by the MIR community [1-5]. Several approaches compute a score or probability *S(t,k)* of being in a key *k* on a given beat *t*. However, these algorithms are not designed to precisely determine where do modulations occur. To better model modulations, we design three proximity measures involved in *S(t,k)*:

- Do the recently heard notes relate to the scale of a key?
- Was there a strong harmonic progression of a key in the recent past of the score?
- What is the most probable key given the key that was used right before?

II. KEY COMPATIBILITY TO THE CURRENT SCALE

Several approaches use histograms to find the key in a score [1,4]. We introduced the *current scale* in [6]. It is built by associating to each of the seven pitch names (e.g. A, B, C,...) the *last encountered accidental* on a given beat. The proximity measure of this current scale to the scale of each key can be estimated as the number of differences between these two scales.

III. HARMONIC PROGRESSION

Quinn [3] states the importance of harmonic progressions to determine keys in a score. The most stable harmonic progression for a key is the one from the dominant (fifth scale degree) to the tonic (first scale degree). When such a progression is encountered on a given beat in the score, the proximity measure should be the lowest possible for the related key and this beat.

IV. KEY RELATIONSHIPS

Given the start key and the musical style, modulations in selected keys are more expected than in other keys. These preferences can be modeled by *pitch spaces* [2,3]. We use the *key relationship table* introduced by Weber [7] to estimate this proximity measure.

V. EXPERIMENT AND RESULTS

S(t, k) is finally computed combining the three measures:

$$S(t,k) = d_{V\to I}(t,k) + d_{diat}(g(t),k) + \min_{k'}[d_{Weber}(k',k) + s(t-1,k')]$$

Finding the best sequence of keys can be computed by dynamic programming.

This model has been tested on a manually annotated corpus of Mozart's string quartets movements[8]. Coefficient values have been estimated on a randomly selected training set. First results indicate that the key is correctly estimated on 90.1 % of the beats of the validation set, ranging from 78.5 % to 98.6 %. The comparison with the existing models is in progress.

REFERENCES

[1] Carol L. Krumhansl. Music psychology : Tonal structures in perception and memory. *Annu. Rev. Psychol.*, 42,, 1991.

[2] Fred Lerdahl. *Tonal Pitch Space*, Oxford University Press, 2001.

[3] Ian Quinn. Are pitch-class profiles really “key for key” ? *Zeitschrift Der Gesellschaft Für Musiktheorie*, 7(2), 2010.

[4] David Temperley. What’s key for key ? The Krumhansl-Schmuckler key-finding algorithm reconsidered. *Music Perception*, 17(1), 1999.

[5] Néstor Napoles Lopes et al.. Key-Finding Based on a Hidden Markov Model and Key Profiles. *DlfM*, 2019

[6] Laurent Feisthauer et al.. Modeling and learning structural breaks in Sonata Form. *ISMIR*, 2019

[7] Jacob Gottfried Weber. *Versuch einer geordneten Theorie des Tonsetzkunst*, 1817-21

[8] Pierre Allegraud et al., Learning Sonata Form Structure on Mozart's String Quartets. *Trans. of the Int. Society for Music Information Retrieval (TISMIR), in press*

Exploring the Aesthetics and Utility of Sonification with Isomorph– Interactive Sonification for Molecular Physics

Joseph Hyde[1*], Thomas J. Mitchell[2*], Helen M. Deeks[3], David R. Glowacki[3], Alex J. Jones[3]

[1*]Bath School of Music and Performing Arts, Bath Spa University, UK, j.hyde@bathspa.ac.uk
[2*]Department of Computer Science and Creative Technologies, University of the West of England, UK
[3*]Intangible Realities Laboratory, School of Chemistry, University of Bristol, UK

Abstract— **This paper presents the Isomorph project, in which we are developing a set of open sonification tools for scientists and others exploring molecular dynamics in realtime simulations. Here we outline our work with a VR-based system, and a study to systematically work towards data-to-sound mappings which are aesthetically satisfying but transparent, ergonomic and demonstrably meaningful.**

I. Background

We discuss our experiences in combining our prototype sonification tools with NarupaXR [1], a system developed at the University of Bristol that allows multiple users to interactively build and manipulate complex molecular structures using VR headsets and controllers.

We have explored a number of applications using the system. These include a drug docking application (Fig. 1), where small ligand molecule - a candidate for a novel drug design - must be positioned within a larger protein molecule This is a complex spatial task that we believe can be greatly aided by sonification.

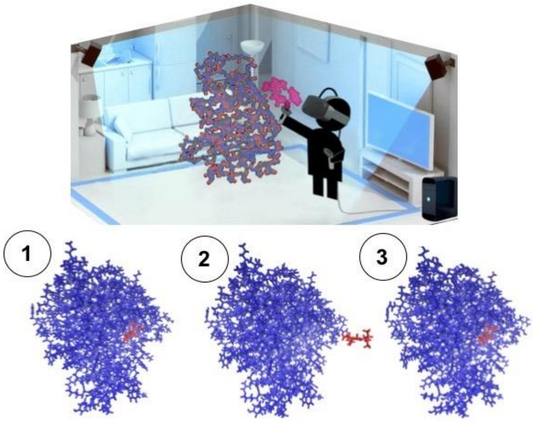


Figure 1. Drug docking experiment: (1) illustrates the ligand (magenta) in its original position, (2) unbound from the protein (purple), and (3) in its final pose after user has docked it to the protein.

II. Sonification

A key focus of Isomorph project is the aesthetics and ergonomics of sonification. We believe that sonification can enrich the apprehension of data in the media-rich environment of VR, but only if fit for purpose. Whilst accurate representation of the data is paramount, sonifiation should also exhibit the qualities of good (sound) design, being perceptually congruent [2], meaningfully interactive and suitable for prolonged use without aural fatigue.

III. Study

In order to explore a variety of sound design approaches, we undertook a two-part controlled study. In the first part, we invited a number of professional sound designers, working in film, TV, gaming, VR and music, to use a variety of sound materials and techniques to sonify key featuress exposed by the Narupa system in a series of auditory representations.

In the second part, these representations formed the basis of a blind/acousmatic test with representative end users. Their responses were measured, and their impressions of the sonification as part of the Narupa simulation were discussed through in-depth interviews.

IV. Findings

In this paper we focus on qualitative data gathered from the end user interviews, which have guided us towards the formation of a prototype aesthetic framework for sonification in with Narupa. We have used this framework to develop a number of sonifications based on the simulation, which we will present to illustrate the paper.

References

[1] O'Connor MB, Bennie SJ, Deeks HM, Jamieson-Binnie A, Jones AJ, Shannon RJ, Walters R, Mitchell TJ, Mulholland AJ, Glowacki DR (2019) Interactive molecular dynamics in virtual reality from quantum chemistry to drug binding: An open-source multi-person framework. The Journal of Chemical Physics 150 (22):220901. doi:10.1063/1.5092590

[2] Ferguson, J. and Brewster, S. (2018) Investigating Perceptual Congruence Between Data and Display Dimensions in Sonification. In: 2018 CHI Conference on Human Factors in Computing Systems, Montréal, QC, Canada, 21-26 Apr 2018, p. 611. ISBN 9781450356206 (doi:10.1145/3173574.3174185)

* Research supported by APEX Grant – Royal Society, Royal Academy of Engineers, British Academy, Leverhulme Trust

DMRN+14: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2019

QUEEN MARY UNIVERSITY OF LONDON
TUE 17 DECEMBER 2019

Alter: An Ensemble Work Composed with and about AI

David De Roure[12*†], Emily Howard[3†], Robert Laidlow[3†‡] and Pip Willcox[14*]

[1]Oxford e-Research Centre, University of Oxford, david.deroure@oerc.ox.ac.uk
[2]The Alan Turing Institute, London, UK
[3]Royal Northern College of Music, Manchester, UK
[4]The National Archives, Kew, UK

***Abstract*— We describe the composition of *Alter*, a chamber ensemble work in which the human composer, Robert Laidlow, worked co-creatively with multiple AI systems. *Alter* was performed in November 2019 as part of "Imagining the Analytical Engine", a musical tribute to Ada Lovelace. Our presentation of this work will include musical examples.**

I. INTRODUCTION

In 2014 we initiated a thought experiment at DMRN: had Charles Babbage built his Analytical Engine, and had Ada Lovelace lived to use it, what music might she have generated? This was a response to Lovelace's note that "the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent." [1] In subsequent years we have explored this question and reported back to DMRN.

This year we have ventured into Artificial Intelligence, inspired by Lovelace's own exploration of the potential of engines to inspire, resemble, assist and enact creativity, as described by Boden. [2] Working with the Centre for Practice & Research in Science & Music (PRiSM) at the Royal Northern College of Music, the work culminated in a premiere of a new chamber ensemble work *Alter* at Milton Court Concert Hall in London on 2nd November 2019. *Alter* was a Barbican commission for the PRiSM team led by Robert Laidlow, PRiSM researcher in AI-Assisted Composition, and was part of an evening of performances under the title "Ada Lovelace: Imagining the Analytical Engine" curated by PRiSM director Emily Howard.

II. COMPOSITION & PERFORMANCE

While AI solutions have become increasingly effective at generating music based on a specific training set, such as the works of one composer, the intent of our work has been not to replace the human composer but rather to assist human creativity. Hence *Alter* is written using artificial intelligence co-creatively, and in fact it is also about artificial intelligence as it explores the very process by which it came about. Through three phases it traces the development of an artificial mind: from hazy, unformed conception to a complex and creative self.

The composition uses AI in multiple ways: sometimes behind the scenes to inform large-scale decisions, sometimes locally where entire phrases are composed by AI music using MuseNet Music Transformer (OpenAI), designed and provided by Christine Payne. The piece also includes electronics: first in recordings of voices that do not exist, produced by DeepMind's WaveNet, and later becoming a true digital counterpart in a human-electronic duet.

The text is written by an AI that audibly develops in coherence and philosophical scope. Coded at The Alan Turing Institute, the AI learns from Ada Lovelace's correspondence, using a language model based on a 19th-century letter corpus supplied by the Electronic Enlightenment team at the Bodleian Libraries. It goes on to use OpenAI's GPT-2, which provides the most extensive training available. Thus the narrative of the scene reflects the data science behind its production.

Alter premiered at Milton Court, performed by members of the Britten Sinfonia. The electronics track provided a disembodied digital counterpart to the real-life mezzo-soprano Marta Fontanals-Simmons. In a further creative response to Lovelace and Babbage, the percussionist used the 'Lovelace Engine', a hand-turned percussion battery styled after Babbage's 19th-century Difference and Analytical Engines. An important part of this project lies in imagining the future that Lovelace and her contemporaries might have seen – one of computers formed of gears, mechanism, industry interacting with people – and juxtaposing that with the reality of AI today.

ACKNOWLEDGMENT

Ursula Martin provided invaluable advice and support. The Lovelace Engine was designed and built by Jonathan Morris with assistance from Cambridge Design Partnership.

REFERENCES

[1] A.A. Lovelace, (Trans.) Sketch of the analytical engine invented by Charles Babbage, with notes by the translator. In Scientific Memoirs, Vol. 3,1843, pp.666-731.

[2] Margaret Boden, *The Creative Mind: Myths and Mechanisms*, London: Weidenfeld and Nicholson, 1990; expanded edition London: Abacus, 1991.

[*] Supported by "Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption", EPSRC award EP/L019981/1.

[†] PRiSM, the RNCM Centre for Practice & Research in Science & Music, is supported by Research England's Expanding Excellence in England Fund.

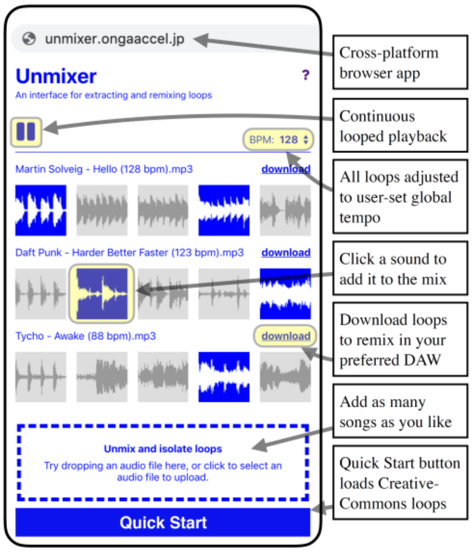
[‡] Supported by the AHRC North West Consortium Doctoral Training Partnership.

Unmixer: An Interface for Extracting and Remixing Loops

Jordan B. L. Smith[1], Yuta Kawasaki[1] and Masataka Goto[1]*

[1]National Institute of Advanced Industrial Science and Technology (AIST), Japan,
unmixer-ml@aist.go.jp

***Abstract*— Someone planning to remix a song would like to have segmented stem tracks at their disposal; that is, isolated instances of the loops and sounds that were used to compose the original song. We present Unmixer, a web service that will analyze and extract loops from any audio uploaded by a user (see Fig. 1). The loops are presented in an interface that allows them to be immediately remixed, and if users upload multiple tracks, they can create mash-ups with the loops, which are automatically matched in tempo.**



unmixer.ongaaccel.jp

Unmixer

An interface for extracting and remixing loops

BPM: 128

Martin Solveig - Hello (128 bpm).mp3 download

Daft Punk - Harder Better Faster (123 bpm).mp3 download

Tycho - Awake (88 bpm).mp3 download

Unmix and isolate loops

Try dropping an audio file here, or click to select an audio file to upload.

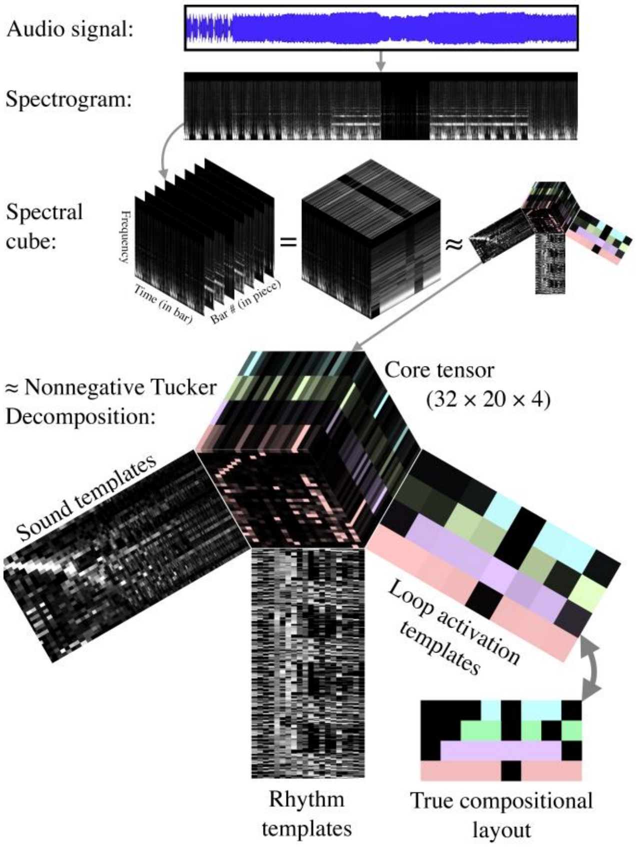
Quick Start

Cross-platform browser app
Continuous looped playback
All loops adjusted to user-set global tempo
Click a sound to add it to the mix
Download loops to remix in your preferred DAW
Add as many songs as you like
Quick Start button loads Creative-Commons loops

Figure 1: Screenshot of Unmixer website annotated with main features.

I. PROJECT DETAILS

To analyze the audio, we adapt a method of source separation that we recently proposed [1], in which a 2D spectrogram is split at each downbeat and stacked into a 3D "spectral cube", allowing us to model periodic repetitions. We estimate the nonnegative Tucker decomposition, which describes the signal very naturally as the product of a set of sounds, rhythms, and loop activations, the latter directly estimating the compositional layout of the estimated loops (see Fig. 2).



Audio signal:
Spectrogram:
Spectral cube:
Frequency
Time (in bar)
Bar # (in piece)
=
≈
≈ Nonnegative Tucker Decomposition:
Core tensor (32 × 20 × 4)
Sound templates
Loop activation templates
Rhythm templates
True compositional layout

Figure 2: Illustration of non-negative Tucker decomposition being applied to a song of length 8 bars, decomposed as a product of 32 sounds, 20 rhythms, and 4 loop activation templates. The loop activation templates reproduce the true composition of this synthetic example.

To reduce the redundancy of some loops, we propose an extra factorization step with a sparseness constraint and demonstrate (in a test using synthesized pieces) that it improves the source separation result. We also propose a method for selecting the best instances of the extracted loops (maximizing their loudness and minimizing cross-talk from other loops) and demonstrate its effectiveness in an evaluation. Both of these improvements are incorporated into the back end of the interface.

REFERENCES

[1] J. B. L. Smith and M. Goto, "Nonnegative tensor factorization for source separation of loops in audio" in *Proceedings of ICASSP*, Calgary, Canada, pp. 171–175.

*This work was supported in part by JST ACCEL Grant Number JPMJAC1602, Japan.

J. B. L. Smith was a Visiting Researcher at AIST Japan when this project was initiated.

Y. Kawasaki is a Creative Engineer in the Information Technology Research Institute at AIST Japan.

M. Goto is Prime Senior Researcher at AIST Japan.
All authors may be reached regarding this project at unmixer-ml@aist.go.jp.

DMRN+14: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2019

QUEEN MARY UNIVERSITY OF LONDON
TUE 17 DECEMBER 2019

The Impact of Dataset Modifications on Music Similarity Measures

Roberto Piassi Passos Bodo[1*], Emmanouil Benetos[2], and Marcelo Queiroz[1]

[1] Institute of Mathematics and Statistics, University of São Paulo, Brazil, {rppbodo, mqz}@ime.usp.br
[2] Centre for Digital Music, Queen Mary University of London, UK, emmanouil.benetos@qmul.ac.uk

I. INTRODUCTION

This project approaches the cover detection problem using music similarity techniques. The problem assumes that there are multiple performances of each musical work, which have to be classified as representing the original version. Thus, it is assumed that some musical property is going to be shared between the audio recordings. A major contribution of our work is the investigation on the impact of modifications in cover song datasets in terms of cover song detection performance.

II. MUSIC SIMILARITY FRAMEWORK

We implemented a music similarity framework that is composed of three modules. The first is the feature extractor, that uses librosa[1] and essentia[2] to compute several useful audio descriptors, such as chromagram, tempogram, pitch contour, etc. The second module is a feature aggregator that is responsible to create a summarized version of the local features for each song. We used basic statistics summarization methods, such as single Gaussian, Gaussian mixtures, as well as other methods such as vector quantization and Markov chains. The last module measures similarity between audio files based on the aggregated feature vectors (which are in the $\mathbb{R}^n$ space), using distance functions such as euclidean, manhattan, chebyshev, etc.

III. DATASET MODIFICATIONS

A. *Frame selection*

Some frames may be harmful to the calculation of similarity between two music recordings because they might not be representative of them. This may occur for instance with low energy frames, and we propose to eliminate a percentile of such frames. In addition to that, we hypothesize that the frames that contain vocals might be more relevant to similarity measures. We therefore decided to select frames also by the energy of the vocals track (obtained using the Open-Unmix [1] source separation algorithm). With these we produce eighteen new secondary datasets (10%, 20%, ... , 90% of the frames in both selection modes).

B. *Segment selection*

The selection of meaningful parts can be extrapolated to whole segments. However, the segments will not be chosen by energy, but instead by their musical content. We selected three musical segmentation algorithms, and three thumbnailing algorithms to produce candidate segments meant to replace the full-length recordings (generating six new datasets). Besides, we also explore selecting the beginning, middle and end portions of the songs with 10s and 30s (generating another six secondary datasets).

C. *Source separation*

Another idea we would like to experiment with is using separated sources within cover song detection. Since we are attacking the problem of music similarity by its individual terms, we can use only the vocal track to obtain melodic similarity measures, only the drums track to obtain rhythmic measures, and so on. For the task of separating sources from the original recordings we are using the Open-Unmix [1] project which is currently the state-of-the-art in audio source separation. We therefore obtain four new datasets, one for each instrumental track (vocals, bass, drums, and other).

IV. PRELIMINARY RESULTS

To this date we have managed to run experiments with the Covers80 dataset [2]. The results displayed in Table I show that every dataset modification here proposed brings an improvement in the Mean Rank (MR) and Mean Reciprocal Rank (MRR) metrics, when compared to using the original dataset. This result encourages us to extend the experiment to other cover songs datasets and dataset modifications.

TABLE I. COVERS80 RESULTS

Best results	***MR***	***MRR***
Original dataset	45.4625	0.178638
Frame selection	42.51875	0.205291
Segment selection	45.06875	0.178040
Source separation	41.98125	0.208763

REFERENCES

[1] F.-R. Stoter and S. Uhlich and A. Liutkus and Y. Mitsufuji, “Open-Unmix - A Reference Implementation for Music Source Separation” in *Journal of Open Source Software*, 2019.

[2] D. P. W. Ellis, “The covers80 cover song data set”, *URL: http://labrosa. ee. columbia. edu/projects/coversongs/covers80, 2017.*

[*] This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001. EB is supported by RAEng Research Fellowship RF/128. MQ is supported by CNPq grant 309645/2016-6.

[1] https://librosa.github.io/

[2] https://essentia.upf.edu/

Dig That Lick: Exploring Patterns in Jazz Solos

Simon Dixon[1*], Polina Proutskova[1], Tillman Weyde[2], Daniel Wolff[2], Martin Pfleiderer[3], Klaus Frieler[3], Frank Höger[3], Hélène-Camille Crayencour[4], Jordan B. L. Smith[4], Geoffroy Peeters[5], Doğaç Başaran[5], Gabriel Solis[6], Lucas Henry[6], Krin Gabbard[7], Andrew Vogel[7]

[1]Centre for Digital Music, Queen Mary University of London, UK, s.e.dixon@qmul.ac.uk
[2]City, University of London, UK [3]University of Music Weimar, Germany [4]National Center for Scientific Research, France [5]IRCAM, France [6]University of Illinois, USA [7]Columbia University, USA

***Abstract*— We give an overview of outcomes from the recently completed project "Dig that lick: Analysing large-scale data for melodic patterns in jazz performances", involving a multi-disciplinary and international team of researchers. On the technical side, the project built infrastructure and tools for extraction, discovery, search and visualisation of melodic patterns and associated metadata. These outcomes facilitate analysis on the musicological side of the use of melodic patterns in improvisation, to answer questions about the origins, evolution and transmission of such patterns. This in turn gives insight into the extent to which improvisers rely on patterns, the development of individual and shared styles, and the level of influence of individual musicians, based on the amount of re-use of their improvised material by later musicians.**

I. BACKGROUND

The recorded legacy of jazz spans a century and provides a vast corpus of data documenting its development. Recent advances in digital signal processing and data analysis technologies enable automatic recognition of musical structures and their linkage through metadata to historical and social context. Automatic data extraction and aggregation give unprecedented access to large collections, fostering new interdisciplinary research opportunities. The *Dig That Lick* (DTL) project developed novel technological and music-analytical methods to gain fresh insight into jazz history.

The importance of musical patterns to jazz is well established in the scholarly literature and popular discourse. Ethnographers analyse how musicians learn and use licks; music psychologists debate the role of licks in improvisation; and fan-generated YouTube videos illustrate the remarkable popularity of one seven-note pattern known simply as *The Lick*. Yet many open questions remain about the development and transmission of licks in individual careers and jazz history. We investigate the usage of patterns and licks in monophonic jazz solos using search algorithms on a large database of jazz solo transcriptions.

II. PROJECT OUTCOMES

The transcriptions are created automatically from commercial audio recordings using state-of-the-art melody extraction algorithms based on neural networks and advanced signal processing methods [1]. New evaluation metrics were developed to assess the suitability of extraction algorithms for tasks such as pattern matching [3]. An overview of the technologies employed in DTL [10], plus a detailed account of the pattern analysis [4] and three case studies [5,7,9] were presented. A study on jazz improvisation was also published [2]. Web-based search applications were developed, allowing the user to search for exact and inexact patterns in multiple melody databases, to specify constraints on the metadata and patterns, and to visualise and hear the retrieved results [6,8]. These interfaces are publicly available at the following web site: https://dig-that-lick.hfm-weimar.de/

REFERENCES

[1] D. Başaran, S. Essid, and G. Peeters. 2018. "Main Melody Estimation with Source-Filter NMF and CRNN". In *19th International Society for Music Information Retrieval Conference.*

[2] Frieler, K. 2019. Constructing Jazz Lines: Taxonomy, Vocabulary, Grammar. In *Jazzforschung heute: Themen, Methoden, Perspektiven*, ed. M. Pfleiderer, W.-G. Zaddach, Berlin: Edition EMVAS, 103-132.

[3] K. Frieler, D. Başaran, F. Höger, H.-C. Crayencour, G. Peeters, and S. Dixon. 2019. "Don't Hide in the Frames: Note- and Pattern-based Evaluation of Automated Melody Extraction Algorithms." In *6th International Conference on Digital Libraries for Musicology*.

[4] K. Frieler, F. Höger and M. Pfleiderer. 2019. "Towards a History of Melodic Patterns in Jazz Performance." In 6*th Rhythm Changes Conf.*

[5] K. Frieler, F. Höger, and M. Pfleiderer. 2019. "Anatomy of a lick. Structure and variants, history and transmission." In *Book of Abstracts of the Digital Humanities Conference, Utrecht.*

[6] K. Frieler, F. Höger, M. Pfleiderer, and S. Dixon. 2018. "Two Web Applications for Exploring Melodic Patterns in Jazz Solos." In *19th International Society for Music Information Retrieval Conference.*

[7] K. Gabbard. 2019, "What We Are Digging Out of the Data." In 6*th Rhythm Changes Conf.*

[8] F. Höger, K. Frieler, M. Pfleiderer, and S. Dixon. 2019. "Dig That Lick: Exploring Melodic Patterns in Jazz Improvisation." *20th International Society for Music Information Retrieval Conference: Late Breaking Demo.*

[9] G. Solis and L. Henry. 2019. "Chasing the Trane: Quantifying the Social Journey of a Coltrane Solo." In 6*th Rhythm Changes Conf.*

[10] T. Weyde, D. Wolff, S. Dixon, P. Proutskova, H.-C. Crayencour, J. Smith, G. Peeters, and D. Başaran. 2019. "Dig That Lick: A Technical Primer for Big Data Jazz Studies." In 6*th Rhythm Changes Conf.*

1* This research was funded under the Trans-Atlantic Program Digging into Data Challenge with the support of the UK Economic and Social Research Council (ES/R004005/1), the French National Research Agency (ANR-16-DATA-0005), the German Research Foundation (PF 669/9-1), and the US National Endowment for the Humanities (NEH-HJ-253587-17).

Author DB is now at Audible Magic, UK, and GP is now at Telecom ParisTech, France.

Modulation Spectra for Musical Dynamics Perception and Retrieval

Luca Marinelli[1], Athanasios Lykartsis[1], and Charalampos Saitis[2]

[1]Audio Communication Group, TU Berlin, Germany
[2]Centre for Digital Music, Queen Mary University of London, UK, c.saitis@qmul.ac.uk

Abstract— **To investigate variations in timbre space with regard to musical dynamics, a convolutional neural network was trained on modulation power spectra of single notes of sustained instruments played at pianissimo and fortissimo dynamics. Samples were rms-normalized to eliminate loudness information and force the network to focus on timbre attributes of dynamics shared across different instrument families.**

I. Introduction

Recent research has shown that even if no loudness cues are available, listeners can still quite reliably identify the intended dynamic strength of a performed sound by relying on timbral features [1]. More recently, acoustical analyses across an extensive set of anechoic recordings of instrument notes played at pianissimo (*pp*) and fortissimo (*ff*) showed that attack slope, spectral skewness, and spectral flatness together explained 72% of the variance in dynamic strength across all instruments, and 89% with an instrument-specific model [2]. The overall aim of the research presented here is to further investigate the role of timbre in musical dynamics, focusing on the contribution of spectral and temporal modulations.

II. Method

Using 33 sustained instruments from the same database as [2], 1 s snippets were extracted from the steady-state part of notes. The modulation power spectrum (MPS) is implemented as the squared amplitude of the two-dimensional Fourier transform of the logarithmic amplitude of the mel-scaled short time Fourier transform (STFT). For each time frame of the STFT, the rms was computed for the whole frequency range and used to normalize the same frame.

The CNN architecture (Table 1) was implemented with Keras running on top of TensorFlow. Average pooling was chosen because max pooling seemed to promote overfitting. All activation functions, but the softmax on the last dense layer, are rectified linear units. Instead of tuning each model separately, a global setup for all experiments was used.

III. Results

The model obtained an accuracy of 91.7% for brass instruments, 97.3% for single reeds, 85.2% for double reeds, 64.9% for bowed strings, and 92.6% in a 10-fold cross validation for the entire dataset.

Through visualization of the *pp* and *ff* saliency maps of the CNN it was possible to identify discriminant regions of the MPS and define an audio descriptor. A linear discriminant analysis with 10-fold cross validation using this MPS-based descriptor on the entire dataset performed better than using two STFT-based spectral descriptors, namely spectral skewness and spectral flatness (43.2% error reduction).

TABLE I.

CNN Architecture	
Layer type	***Parameters***
Conv2D	filters: 16, size: 7x7, stride: 3
Batch norm.	--
Conv2D	filters: 32, size: 3x3, stride: 1
Batch norm.	--
Average pool.	size: 2x2
Conv2D	filters: 64, size: 3x3, stride: 1
Average pool.	size: 2x2
Flatten	--
Dense	neurons: 128
Dropout	p: 0.5
Dense	neurons: 2

Overall, audio descriptors based on different regions of the MPS could serve as sound representation for machine listening applications, as well as to better delineate the acoustic ingredients of different aspects of timbre perception. Future work should expand on impulsive sounds and include different dynamic gradations.

References

[1] M. Fabiani and A. Friberg, "Influence of pitch, loudness, and timbre on the perception of instrument dynamics" *J. Acoust. Soc. Am.*, 130(4), 2011, pp. EL193–EL199.

[2] S. Weinzierl, S. Lepa, F. Schultz, E. Detzner, H. von Coler, and G. Behler, "Sound power and timbre as cues for the dynamic strength of orchestral instruments" *J. Acoust. Soc. Am.*, 144(3), 2018, pp. 1347–1355.

Searching for Efficient Processing Pipelines Applied to MIR in Embedded Systems

Filipe Lins[1*], Marcelo Johann[2] and Rodrigo Schramm[3]

[1*,2] Instituto de Informática, UFRGS, Brazil
[3]Departamento de Música, UFRGS, Brazil, rschramm@ufrgs.br

***Abstract*— This work aims to find efficient computational pipelines applied to Audio Signal Processing and Music Information Retrieval (MIR) algorithms running on embedded systems. In order to support the design choices, we perform benchmark experiments and measure the power estimation, execution time and resource utilization. Our preliminary results are evaluated using the System on Chip (SoC) Zynq-7000, which has a dual-core ARM Cortex-A9 processor with a Field Programmable Gate Array (FPGA) logic fabric on the same chip[1].**

I. Introduction

The Internet of Things (IoT) combines several technologies into small hardware devices and opens a new range of possibilities in the field of Audio Signal Processing and Music Information Retrieval - MIR. A common requirement for MIR algorithms is real-time processing and low latency [2, 3, 4], in which the computing demand is a challenge for small embedded devices and often comes at the price of high power usage. This research aims to design a processing pipeline that explores optimal hardware and software configurations on available technologies, including Microcontrollers, Microprocessors, Systems on Chip (SoC), Graphics Processing Unit (GPUs), Field Programmable Gate Array (FPGAs), and Application Specific Integrated Circuits (ASICs).

In order to evaluate the impact of MIR algorithms regarding power consumption and real-time processing on embedded systems, we designed a set of experiments based on common algorithms used in performance benchmarks, including FFT and convolution. There are many different configurations inside these hardware platforms that need to be tested since they might directly affect the execution time and resource utilization such as different implementations of FFT, the number of channels, input data and phase factor width, data format, scaling options and rounding modes. Besides these configurations, there are also the design choices of hardware architectures such as the usage of Block Random Access Memory versus Double Data Rate (DDR), custom vector floating point registers, and Direct Memory Address (DMA).

II. Experiments

We started this benchmark using two main hardware configurations in the Zynq-7000 architecture [1], using complex data stored in separate real/image arrays. The first chosen configuration on Zynq-7000 was a dual-core ARM-A9, storing the data in DDR memory with different compiler options related to Neon Registers and Vector Floating Point (VFP) extension. The second configuration was a dual-core ARM-A9, storing the data in DDR, but the FFT algorithm is running on the FPGA using the Xilinx® LogiCORE™ Intellectual Property (IP). This core implements the Cooley-Tukey FFT algorithm and accesses the data using DMA. Comparisons among different configurations show a high variability in the execution times and power consumption.

III. Final Considerations

At this first stage of the research project, we are developing the pipeline that defines the set of hardware and parameter configurations that will be further used to evaluate all the resources of these systems when implementing complex MIR algorithms. The next stage of this research aims to evaluate algorithms for convolution, spectrogram representation (Short Time Fourier Transform - STFT, Constant-Q Transform - CQT), iterative estimations (Non-negative Matrix Factorization - NMF, Probabilistic Latent Component Analysis - PLCA), and complex machine learning routines (Random Forest, ANN).

References

[1] CROCKETT, Louise H. et al. *The Zynq Book: Embedded Processing with the Arm Cortex-A9 on the Xilinx Zynq-7000*. All Programmable Soc. Strathclyde Academic Media, 2014.

[2] STARK, Adam; PLUMBLEY, Mark. *Real-time chord recognition for live performance.* In: Proceedings of ICMC 2009. p. 85-88, 2019.

[3] VACA, Kevin; GAJJAR, Archit; YANG, Xiaokun. *Real-Time Automatic Music Transcription (AMT) with Zync FPGA.* In: Proceedings of ISVLSI/ IEEE, 2019. p. 378-384. 2019.

[4] SCHRAMM, Rodrigo; VISI, Federico; BRASIL, André; JOHANN, Marcelo. *A polyphonic pitch tracking embedded system for rapid instrument augmentation.* In: Proceedings of NIME 2018, pp. 120–125, 2018.

*This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

MSc F. Lins is a PhD candidate in the Institute of Informatics, UFRGS/Brazil. (email: fmlins@inf.ufrgs.br)

Dr. M. Johann is an Associate Professor in the Institute of Informatics, UFRGS/Brazil. (email: johann@inf.ufrgs.br)

Dr. R. Schramm is a Lecturer in the Music Department, UFRGS/Brazil. (corresponding author e-mail: rschramm@ufrgs.br).

Homepage and Search Personalization at Spotify

Mi Tian[1], Rishabh Mehrotra[1], Lucas Maystre[1], and Mounia Lalmas[1]

[1]Spotify, UK, mtian@spotify.com

***Abstract*— Spotify's personalized services provide users access to a massive repository of audio content. This talk gives an overview of Spotify research on homepage and search personalization, experimentation and evaluation, as well as our pursuit of a bias-free machine learning practice.**

I. Spotify as a personalized audio assistant

Music and podcast listening is a commonplace activity in people's everyday life. With the advent of online music streaming services, listeners have access to an ever-growing catalog of music and audio content. A big challenge for users is to effectively retrieve what they want to listen to and discover what they may enjoy listening. At Spotify, our mission is to match listeners and artists in a relevant and personalized way. The main tools to achieve this are Home and Search.

II. Personalization via homepage and search

Home recommendation and *search* are two sides of the same coin at Spotify [1]. They work together to help users get the music and podcast content they will enjoy listening.

Home is a "push" paradigm. It is the default landing page and surfaces the best Spotify has to offer, accommodating both *passive listening* and *active engaging* user mindsets. Home surfaces personalized, context-aware and explainable recommendations via a contextual bandit algorithm [2], which leverages user's streaming behavior to estimate the reward function used to train the bandit model [3]. Personalization is based on contextual features describing what the system knows about the user, their context, and the audio content. Model development and iteration is guided by large scale evaluation through both offline counterfactual evaluation using randomized log data and online A/B tests [2,5]. Furthermore, experimental evaluation highlights the systems is capable of jointly optimizing for user engagement objectives as well as artist preferences [4].

Search is a "pull" experience. User search mindsets are modeled as *focused* or *open and exploratory* based on the query real-time [6]. For each user mindset and expectation we provide a ranked list of search results leveraging between relevance to the query and search recommendations to enable exploration. To build a model that optimizes for various search intentions, we define a set of success signals using the *mixed methods methodology* [7, 8]. From this, we develop a composite *search success* metric at the session level: a session is deemed successful when any one of the success signals is present. We compared our proposed search success rate metric against click-through rate with an online experiment. Both metrics agree on the directionality (they detect the same positive treatment effect), but search success rate is more sensitive than click-through rate [8].

III. Personalization that cares, recommendation that sustains

Individual and cultural representation in music and entertainment have an enormous social impact. Spotify pays close attention to understand the critical role our platform and our algorithms play in ensuring fairness and a level playing field. While heavily relying on data-driven approaches for personalization and recommendation, we also research into guarding against the potential algorithmic bias with various machine learning practices including rebalancing of training data and localized algorithm design and evaluation [4,9]. Besides measuring the short-term user satisfaction of our models, we also monitor their long term impact in broader context such as discovery of new and unrepresented artists as well as playlist diversity.

References

[1] N. J. Belkin, W. B. Croft "Information filtering and information retrieval: Two sides of the same coin", Communications of the ACM 1992

[2] J. McInerney, B. Lacker, S. Hansen, K Higley, H.Bouchard, A. Gruson and R. Mehrotra, "Explore, exploit, explain: Personalizing explainable recommendations with bandits", RecSys 2018

[3] P. Dragon and R Mehrotra and M. Lalmas, "Deriving user- and content-specific rewards for contextual bandits", WWW 2019

[4] R. Mehrotra, J. McInerney, H. Bouchard, M. Lalmas and F. Diaz, "Towards a Fair Marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems", CIKM 2018

[5] A. Gruson, P. Chandar, C. Charbuillet, J. McInerney, S Hansen, D. Tardieu and B. Carterette, "Offline evaluation to make decisions about playlist recommendation algorithms", WSDM 2019

[6] A. Li, J. Thom, P. Ravichandran, C. Hosey, B. St. Thomas andJ. Garcia-Gathright, "Search mindsets: Understanding focused and non-focused information seeking in music search", WWW 2019

[7] C. Hosey, L.Vujović, B. St. Thomas, J. Garcia-Gathright and J. Thom, "Just give me what I want: How people use and evaluate music search", CHI 2019

[8] P. Chandar, J. Garcia-Gathright, C. Hosey, B. St. Thomas and J. Thom, "Developing evaluation metrics for instant search using mixed methods", SIGIR 2019

[9] H. Cramer, J. Garcia-Gathright, S. Reddy, A. Epps, A. Springer and R. T. Bouyer, "Translating, tracks and data: An algorithmic bias effort practice", CHI 2019

Automatic Music Accompaniment with a Chroma-based Music Data Representation

Lele Liu[1] and Emmanouil Benetos[1]

[1]School of Electronic Engineering and Computer Science, Queen Mary University of London, UK,
lele.liu@qmul.ac.uk

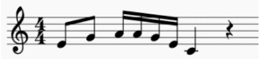
Abstract— **We propose a model to generate melodic music accompaniment for classical piano music using Long-Short Time Memory (LSTM) networks. We design a chroma-based music data representation that combines music knowledge including pitch chroma, pitch height, onsets, tempo, and modes. A subjective listening test shows that by using the musical features in the music data representation, the LSTM model can generate better accompaniment compared to using simple MIDI pitch numbers in the data representation.**

I. INTRODUCTION

Automatic music generation, aiming at generating music using machine algorithms, has attracted a lot of interest in recent years due to machine and in particular deep learning. Many music generation systems use MIDI-like encodings [1]. We pay special attention to how different musical features can influence the performance of the music data representation. By taking musical features into account, we describe pitch by pitch chroma and pitch height. This operation greatly decreases the dimensionality of the data representation. We also include beat starts and music modes in the input features. This data representation allows the LSTM network to learn easier and generate more harmonious music accompaniments compared to a MIDI-pitch representation. A subjective user study shows that the musical features (especially the chroma feature) used in our proposed data representation increase the users' overall evaluation on the music accompaniments.

II. METHODOLOGY

Although it is common to use MIDI pitch numbers in music data representations, people perceive pitch as having two dimensions, where pitch chroma plays a role in representing melodies and pitch height helps in separating music streams and sound sources [2]. Thus, we use chroma features and separate MIDI pitch into 12 pitch chromas and 11 pitch heights. This encoding describes pitch in a more musical way and decreases the dimension of a pitch encoding from 128 to 23. We divide the data representation for a monophonic music piece into two parallel sequences -- a melody sequence *M* and a music annotation sequence *A*. We build the *M sequence* using note pitch chroma, pitch height, note sustain and rest, and the *A sequence* using beat information and music mode. The two sequences are sampled in parallel along the music pieces based on a musically relevant time step. An example of encoding the *M sequence* for a monophonic music piece is shown in Figure 1.



Symbolic *M sequence*: [E4 - G4 - A4 A4 G4 E4 C4 - - - 0 0 0 0]

Figure 1. Encoding *M sequence* for a monophonic music piece.

We assume that all music pieces are within the style of Western tonal music in major/minor modes and contain a monophonic melody and a monophonic accompaniment. The model is based on a recurrent neural network architecture, generating music accompaniment from left to right and predicting one note (or rest/sustain) at each musically relevant time step. The model is composed of two components - a *pitch chroma component* and a *pitch height component*. Each component contains one embedding layer, one LSTM layer and a SoftMax activation output layer. We define both components to solve a categorical classification problem. The final system predicts the accompaniment pitch chroma part at first and uses the chroma part result to predict pitch height.

III. EXPERIMENT RESULTS

We experiment on different data representations within the LSTM network structure and evaluated their performance in a user study. In the user study, 20 listeners (average 5.85 years of musical training) are asked to rate the overall quality of the generated accompaniment. The statistical data of the listeners' ratings are shown in Table 1. According to the ratings, our proposed data representation (No.5) showed the best performance. The t-values and p-values are calculated from t-tests between the specific data representation and the MIDI-pitch data representation.

TABLE I. STATISTICAL DATA OF PARTICIPANTS' RATINGS.

No	*Data Rep*	*Mean*	*Median*	*t-Value*	*p-Value*
1	MIDI-pitch	5.698	6	--	--
2	chroma-only	6.368	**7**	2.278	**0.02384**
3	chroma+beat	5.823	6	0.409	0.68271
4	chroma+mode	5.674	6	-0.078	0.93807
5	proposed	**6.847**	**7**	3.872	**0.00015**

REFERENCES

[1] F. T. Liang et al., Automatic stylistic composition of bach chorales with deep lstm. In ISMIR, pages 449–456, 2017

[2] J. D. Warren et al., Separating pitch chroma and pitch height in the human brain. Proceedings of the National Academy of Sciences of the United States of America, 100 17:10038–42, 2003.W.-K. Smith, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

A MELD TimeMachine for Wagner's Lohengrin

David Lewis[1], Kevin Page[1] and Laurence Dreyfus[2]

[1]Oxford e-Research Centre, University of Oxford, UK, david.lewis@oerc.ox.ac.uk
[2]Faculty of Music, University of Oxford, UK

***Abstract*— We demonstrate a web application optimized for touch interfaces that supports the musicological exploration of an opera, supporting textual and video essays.**

I. SHARING MUSICOLOGY

Musicological argument is frequently concerned with diverse subjects and evidential materials, each potentially exemplified by different media, and each a springing point for further exploration of the author's argument. Despite this, it has traditionally been communicated in linear, textual writings, illustrated by occasional figures. While exploring the referenced materials may be non-linear, this is neither embodied in nor enabled by the communication mediums.

We introduce a tablet-based interactive application which presents a comprehensive digital exploration as a companion to a complete musicological article about the Wagner opera Lohengrin. The article and digital companion show how one motive is altered each time it recurs in the opera, reflecting its role in the drama.

II. USING LINKED DATA WITH MELD

Musicological analysis is encoded – along with relationships to multimedia materials – using Linked Data as an independent, repurposable, and open Research Object. Interactive user views are generated dynamically in the browser directly from this knowledge graph using novel visualisations, enabling the user to navigate all possible paths through the evidential multimodal materials.

The application is built with a new version of the MELD (Music Encoding and Linked Data) framework. MELD traverses Linked Data graphs to select and filter relevant information, with reusable components for creating and retrieving annotations, and for displaying and interacting with musical, textual, graphical and audio-visual materials. MELD is written in Javascript and Python, with resources using standards including the MEI music encoding, TEI, the Music Ontology and Web Annotations.

III. THE APP

Our companion shows the different compositional devices Wagner uses to vary his motives, browsing the whole opera for motive occurrences and their musical and textual contexts. Visualisations and recordings support the analysis, making it more accessible to a general audience. Exploration of relevant materials can follow or be triggered by the article, but can also be reader-driven, with free browsing of the curated musical landscape. A video essay also provides a source of narrative paths through the companion, as a guide itself and as a source of starting points.

We provide two views for music notational content. Vocal score reductions are rendered from MEI with structural analysis dynamically overlaid; annotations trigger audio playback from that point. A second notational visualisation of MEI simplifies the full Wagnerian orchestral score: each instrument playing at a particular time is shown as a coloured ribbon, with the instrument's section of the orchestra providing the colour. This highlights differences of instrumental colour that may be invisible in a vocal score.

For an opera thousands of bars long, overviews are crucial. An ever-present timeline shows all occurrences of a motive, providing a visual summary and a base for navigation. In the Time Machine view, users can also flick through motive occurrences – visualised as libretto, vocal score or orchestration – summarising the sequence within the opera, supporting quick comparisons, and as an index to detail views.

ACKNOWLEDGMENT

The Lohengrin TimeMachine application acts as a companion to the essay "Asking a Forbidden Question: A new view of musical themes in Wagner's Lohengrin" by Prof. Laurence Dreyfus, and builds upon research with important contributions from Carolin Rindfleisch. David Weigl is lead developer of the MELD framework. We thank Will Elsom for his work on the Lohengrin Time Machine design specification, Steve Pierce and Oxford Filmshed Productions for the video tutorial, and Ralph Woodward for his music engravings.

* The Lohengrin TimeMachine was developed by the AHRC-funded Unlocking Musicology project (AH/R004803/1). Unlocking Musicology builds upon research undertaken during the Transforming Musicology project (AH/L006820/1). The MELD framework was created as part of the Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption project, funded by the UK Engineering and Physical Sciences Research Council grant EP/L019981/1.

RadioMe: Artificially Intelligent Radio for People with Dementia

Satvik Venkatesh,[1] David Moffat[1] and Eduardo Reck Miranda[1]
[1*]Interdisplinary Centre for Computer Music Research, University of Plymouth, UK,
satvik.venkatesh@plymouth.ac.uk, david.moffat@plymouth.ac.uk

***Abstract*— RadioMe aims to perform real-time radio remixing for people with moderate to mild dementia. It would devise a way for them to live independently at homes and reduce the need for carers and family members. Research has suggested that music can help patients regulate emotions such as depression, aggression, and anxiety. Therefore, this project would control agitation by playing back music based on their mood. Additionally, it addresses symptoms like memory loss by providing them with personalised reminders to perform daily tasks.**

I. Introduction

Dementia is a syndrome that hampers the regular functioning of the brain. It includes symptoms such as memory loss, difficulties with performing daily tasks, and problems with thinking speed and judgement. Dementia generally affects individuals after the age of 65 years and it is estimated that around 1 million people will suffer from the syndrome by 2025.

Among the elderly population, it is common for people to listen to the radio for entertainment. Additionally, many suggest that it prevents them from feeling lonely. RadioMe aims to remix and customise radio programmes to assist people suffering from dementia. This would enable such individuals to prolong their independent stay at homes before going to care homes and therefore, minimising the need for carers and the responsibility on family members.

RadioMe includes collaborators from different universities, each of them working on different aspects of the project. Specifically, our work package includes audio classification of radio signals and speech synthesis of a DJ-like voice that assists the individuals with their daily tasks.

II. Audio Classification

This project aims to develop a system aimed at assisting dementia patients, with managing agitation episodes. Audio classification will be necessary to parse and understand the current audio output, to understand in which and what way this broadcast can be interrupted. The radio programme should be segregated into audio classes like news, weather report, music, and sports, to name but a few [1]. Furthermore, a song can be interrupted and replaced with a more relaxing song in order to regulate the mood and emotions of the individual.

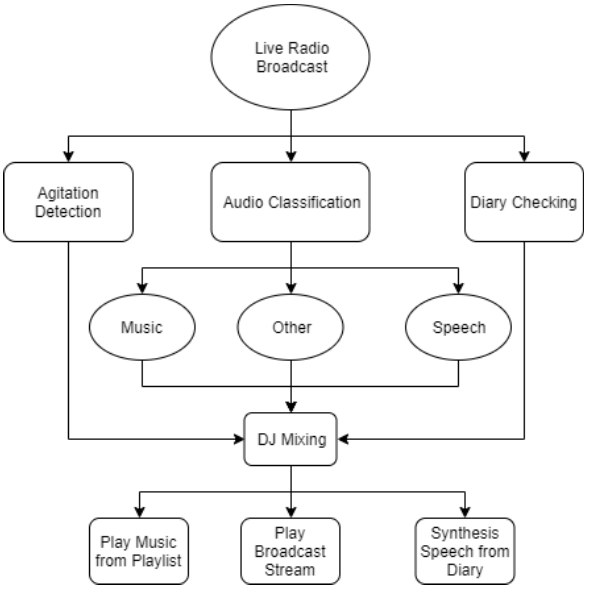


Figure 1. Flow diagram of the RadioMe Core Audio System.

III. Audio DJ-Mixing

To allow a seamless transition between existing radio content, and the personalised experience of RadioMe, there will need to be some subtle radio content mixing and transitions. This will be informed by the audio classification step, and the type of audio content being transitioned to [2]. The capacity to seamlessly transition between different types of audio content, is vital to ensure that further confusion and agitation is not caused.

IV. Personalised Diary Reminders

The added benefit of a personalised radio, is that diarised reminders can be created, to aid with memory. This needs to be used in a supporting manner, through careful phrasing such as “Have you had a nice lunch today?” rather than saying “it is lunch time”. These diary reminders will be overlayed the start of music tracks, or replace the radio DJ usual speech. The use of speech synthesis and replication techniques, including those of recent *deep fakes*, could ensure that the synthesised voice sounds exactly like the current radio DJ.

References

[1] Dhanalakshmi, P., S. Palanivel, and Vennila Ramalingam. "Classification of audio signals using SVM and RBFNN." Expert systems with applications 36.3 (2009): 6069-6075.

[2] Thalmann, Florian, Lucas Thompson, Mark Sandler, and others. 2018. “A User-Adaptive Automated DJ Web App with Object-Based Audio and Crowd-Sourced Decision Trees.” In *4th Web Audio Conference (Wac)*. Berlin, Germany.

*Research supported by Engineering and Physical Sciences Research Council (EPSRC) grant EP/S026991/1.

All authors are with the Interdisciplinary Centre for Computer Music Research (ICCMR), University of Plymouth, UK.
Satvik Venkatesh: satvik.venkatesh@plymouth.ac.uk
David Moffat: david.moffat@plymouth.ac.uk
Eduardo Miranda: eduardo.miranda@plymouth.ac.uk

A Decentralized MELD Agent Framework for Computational Analysis of the Live Music Archive

Graham Klyne[1*], Thomas Wilmering[2], John Pybus[1], and Kevin Page[1]

[1*]Oxford e-Research Centre, University of Oxford, UK, graham.klyne@oerc.ox.ac.uk
[2]Queen Mary University of London, UK

***Abstract*— This poster describes work to create a MELD implementation of Live Music Archive computational analysis.**

I. Background

MELD [7] allows various music-related media to be connected/synchronized with each other via musically meaningful structure, independently of vagaries such as timing within particular performances. SOFA [2] is a tool originally developed for guided music composition from semantically described music fragments. We have been using ideas from SOFA to re-implement the logic of rCALMA [1], a tool for displaying musicological analyses over multiple performances of a work, as a MELD application.

II. Implementation

SOFA [2] operates on a workset of broadly homogeneous elements with semantic annotations (***LDP containers***), and employs multiple independent analyses of workset elements (***agents***) to create new containers of annotations that are used to inform selection options and to provide values for display, and a web application that reads these annotations to discover selection options, and build a corresponding user interface (***REST, uniform interface, and hypermedia-based discovery***). The resulting display is in the form of a grid, whose visible contents are rendered by independent agents. Element descriptions use existing generic linked data vocabularies, such as MELD [9], FRBR [3], MO [4], etc. Analysis is separated from presentation, so an application can present new analyses as agents are implemented.

Linked data standards re-used are Web Annotations [5], and Linked Data Platform [6], which are layered on RDF and HTTP respectively. Our server-side elements use an existing software implementation, with no special server applications. By building on these standards, the (meta)data created is also available for repurposing by other applications.

III. Future Work

New agents to support new forms of analysis and presentation within the existing framework These would be self-contained, not requiring any changes to existing software components, unless some new capability is required that doesn't fit the workset and grid model.

Orchestrating agent activation: experiments with LDP container notifications indicate that we can arrange for agents to be activated automatically when data is updated given some means to connect agents to data elements, probably involving some additional linked data hypermedia structures, which have not yet been designed.

Optimization: the current implementation makes no concessions to efficiency. By building on a uniform interface (LDP), we can identify common access patterns and create optimized paths (in the form of specialized LDP implementations) to overcome inefficiencies in the current implementation.

Acknowledgment

This work is supported by Fusing Semantic and Audio Technologies for Intelligent Music Production and Consumption funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/L019981/1, a collaboration between Queen Mary University of London, University of Nottingham and University of Oxford.

References

[1] Kevin R Page, Sean Bechhofer, György Fazekas, David M Weigl, and Thomas Wilmering. 2017. Realising a layered digital library: exploration and analysis of the live music archive through linked data. In Proceedings of the 17th ACM/IEEE Joint Conference on Digital Libraries (JCDL '17). IEEE Press, Piscataway, NJ, USA, 89-98.

[2] De Roure, D., Klyne, G., Pybus, J., Weigl, D. M., & Page, K. (2018). Music SOFA: An architecture for semantically informed recomposition of Digital Music Objects (pp. 33–41). Association for Computing Machinery.

[3] Functional Requirements for Bibliographic Records (FRBR), https://www.ifla.org/publications/functional-requirements-for-bibliographic-records

[4] Music Ontology, http://purl.org/ontology/mo/

[5] Web Annotation Vocabulary, W3C Recommendation 23 February 2017, https://www.w3.org/TR/annotation-vocab/

[6] Linked Data Platform 1.0 (LDP), W3C Recommendation 26 February 2015, http://www.w3.org/TR/ldp/

[7] Weigl, D., & Page, K. (2017). A framework for distributed semantic annotation of musical score: "Take it to the bridge!". International Society for Music Information Retrieval.

AMT for Musicians: Performed-MIDI-to-Score Transcription

Francesco Foscarin[1*], Florent Jaquemard[2,3], Philippe Rigaux[2] and Raphaël Fournier-S'niehotta[2]

[1*]CNAM, Paris, francesco.foscarin@cnam.fr
[2]CNAM, Paris
[3]INRIA, Paris

***Abstract*— Most of the research in automatic music transcription produces signal-oriented outputs, convenient for computer tasks but not for human interpretation and interaction. We advocate the production of music scores, that convey much more meaning to a musician. Our work addresses specific challenges related to score production in the context of classical piano, performed-MIDI-to-score transcription.**

I. INTRODUCTION

Automatic music transcription is the action of converting a music performance into a high-level representation of its musical content. Many definitions of the problem are given in the literature, according to the input and output representations. Among them there are audio file, unquantized piano-roll, quantized piano-roll and music score (the last three being grouped by the name of symbolic music).

In the literature only few works target the production of music scores, while most of it focus on audio to piano-roll transcription. A piano-roll is a good representation for most typical symbolic MIR tasks, *e.g.* melodic similarity, chord extraction, music search and automatic accompaniment. On the other side, the music score excels in being readable by humans and conveying semantic level information (*e.g.* beaming, tuplets, pitch spelling, staves division and tonality).

With the intent of producing a notation exploitable by musician, we need systems that produce a music score as output. The latest works in this direction targets 4 voices polyphonic music [1] (strings quartets and Bach chorales) and piano classical music [2]. The transcription model in the latter work is based on heuristics [3], but the state of the art for many sub-problems has shifted towards machine learning models and we believe it's time to employ those techniques for a full transcription procedure.

II. GOALS AND CHALLENGES

We target piano classical music starting from performed MIDI files (a format produced from a MIDI keyboard, close to piano-roll) with the goal of producing a music score. Targeting piano music presents some interesting challenges. Compared to multi-instrumental music, the piano-roll matrix is sparser and without clear rhythmic patterns. This, paired with the severe use of expressive tempo (*e.g.* rallentando or rubato), drastically reduce the performances of state-of-the art approaches. Additionally, the datasets of annotated piano music are extremely small compared to other kind of music. This has led to prefer heuristics methods over ML techniques in the past, but the situation has changed recently, thanks to some datasets released by people working in modeling expressive piano performances [4]. The last challenges are the separation of different voices and the production of a music score, a difficult format to handle by machine because of its hierarchical structure, whose rich semiology is mostly unaddressed by formal studies.

III. OUR PROPOSAL

A piano MIDI performance to music score framework must perform the following tasks: tempo detection, beat and downbeat tracking, voices separation, metric detection, rhythmic quantization and score structuring. We propose a model based on grammars [5] (Figure 1) that can be trained on music scores [6] and jointly handles rhythmic quantization, metric detection and score structuring. We are currently studying approaches to merge the current results with the output of other models, to supply a fully automated transcription framework.

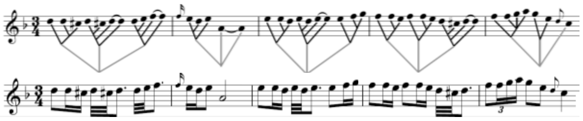


Figure 1. Example of beaming structuring for score production.

REFERENCES

[1] M. A. Román, A. Pertusa, and J. Calvo-Zaragoza, "A holistic approach to polyphonic music transcription with neural networks". In *Proc. Int. Society Music Information Retrieval Conf.*, 2019.

[2] A. Cogliati, D. Temperley, and Z. Duan, "Transcribing human piano performances into music notation". In *Int. Society Music Information Retrieval Conf.*, 2016, pp. 758–764.

[3] D. Temperley, "A unified probabilistic model for polyphonic music analysis". *Journal of New Music Research*, 2009, 38(1):3–18

[4] D. Jeong, T. Kwon, Y. Kim, K. Lee, and J. Nam, "VirtuosoNet: A hierarchical RNN-based system for modeling expressive piano performance". In *Int. Society Music Information Retrieval Conf.*, 2019.

[5] F. Foscarin, F. Jacquemard, P. Rigaux, and M. Sakai, "A Parse-based Framework for Coupled Rhythm Quantization and Score Structuring". In *International Conference on Mathematics and Computation in Music*, 2019, pp. 248-260.

[6] F. Foscarin, F. Jacquemard, and P. Rigaux, "Modeling and Learning Rhythm Structure". In *Sound and Music Computing Conference*, 2019.

*Research supported by ANR France-Quebec MUNIR Project.

DMRN+14: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2019

QUEEN MARY UNIVERSITY OF LONDON
TUE 17 DECEMBER 2019

Using Different Feature Selection Methods for Mood Prediction

Cornelia Metzig[1*] and Mark Sandler[1]

[1*]Centre for Digital Music, Queen Mary University of London, UK, c.metzig@qmul.ac.uk

***Abstract*— We study how the mood variables valence and arousal of a song can be described with extracted features. We study several methods to reduce the feature space.**

I. METHOD

Mood prediction of music has been addressed extensively [1-2]. A large number of low and high features are available to study it, which we do here with vamp plugins. We use them to predict valence and arousal of the moodplay dataset. Many of the plugins give time series, from which we calculate the following summary statistics: mean, standard deviation, skewness, kurtosis, autocorrelation and entropy. Every song can be represented as a vector in these dimensions. Using high dimensional data can be powerful, but comes with the caveat of aggregation artifacts like 'hubness' (type I errors in classification tasks). Out of 500 dimensions, we want to select those that are most informative for mood prediction.

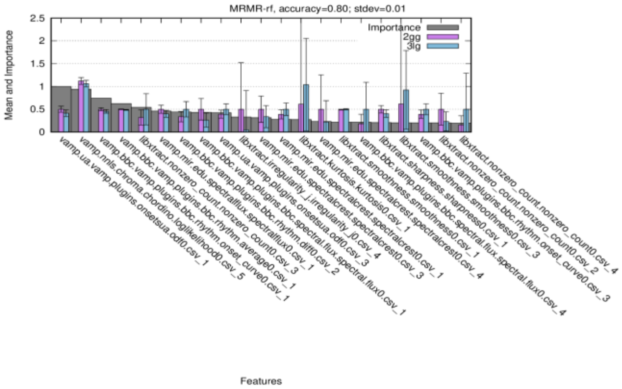


Figure 1: Example of features selected for valence prediction. Grey: Importance of Random forest; purple: songs with high valence; blue: songs with low valence.

One method is (i) to use feature selection of random forest classification (see Fig. 1). Other methods are to calculate summary statistics of all feature across the dataset (see Fig. 2), like (ii) entropy, (iii) coefficient of correlation, or (iv) standard deviation, and to use the features with the highest values. We compare the distances of songs in terms of these features to the distances in terms of valence and arousal distance from human annotations.

II. APPLICATION

We implemented one algorithm for the moodplay website **moodplay.github.io** to make it more interactive. Users can upload songs share them with their community. The song is then placed on the moodplay [3] plane according to the following algorithm:

– from the song uploaded by the user, all vamp plugins and summary statistics are calculated;

– the neighbor songs are identified to which it has the shortest distance;

– it is placed on the mood plane in the middle of these similar songs.

Users can then compare where they would place the song on the valence-arousal plane themselves.

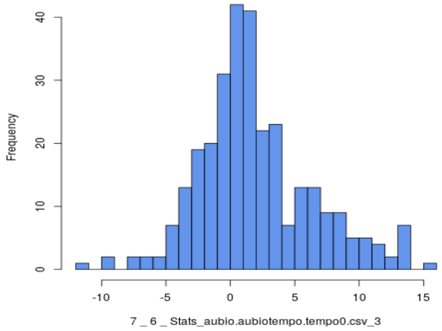


Figure 2: Skewness of aubiotempo, a feature with high entropy distribution.

REFERENCES

[1] Hu, Xiao, Kahyun Choi, and J. Stephen Downie, "A framework for evaluating multimodal music mood classification", *Journal of the Association for Information Science and Technology* 68.2 (2017): 273-285.

[2] Barthet, M., Fazekas, G., & Sandler, M. (2013). "Music emotion recognition: From content-to context-based models". In M. Aramaki, M. Barthet,R. Kronland-Martinet, & S. Ystad (Eds.), From sounds to music and emotions (pp. 228–252). Berlin, Heide

[3] "Join my party! How can we enhance social interactions in music streaming", Allik, A., Thalmann, F., Metzig, C., Sandler, M., *Web Audio Conference, Dec* 4-6, Trondheim, NO

The Algorithmically Enhanced Stylophone

David De Roure[12*], Alan Chamberlain[3*], Iain Emsley[4], John Pybus[1*]
[1]Oxford e-Research Centre, University of Oxford, UK, david.deroure@oerc.ox.ac.uk
[2]The Alan Turing Institute, London, UK
[3]University of Nottingham, UK
[4]School of Media, Film and Music, University of Sussex, Falmer, UK

Abstract— **We describe the development of algorithmically enhanced Stylophones and their use in public engagement events to stimulate conversations about music, mathematics and algorithms.**

I. INTRODUCTION

Over the past 5 years we have pursued a series of developments under the theme of 'numbers into notes' which have been used in research and in public engagement. These include programming an emulator for Babbage's Analytical Engine, a web application (http://numbersintonotes.net), Arduino-based interventions, a 'Fibonacci Theremin', 19th century mathematical algorithms in Logic Pro, and multiple creative works by Alan Chamberlain. [1]

In some cases we have experimented with entirely new interfaces for musical expression, but the motivation for the Logic Pro example was to bring the algorithms to a standard piano keyboard interface while giving the sense of operating algorithms—rather like interacting with a large mechanical 19th century machine. Akin to an arpeggiator, pressing a key instead initiates a mathematical sequence of notes.

The transition from MIDI keyboard to Stylophone was prompted by the desire to engage with a family audience of all ages and musical backgrounds, firstly at an event at the Tate Modern ('Living with the Internet of Things', February 2019) and then in Oxford (Oxford Ideas Festival, October 2019). Here we describe our experiments with these 'algorithmically enhanced Stylophones', and their use to stimulate conversations about music, mathematics and algorithms.

II. THE STYLOPHONE

The Stylophone was first manufactured in the late 1960s by Dubreq Ltd. Although variants are available, the basic concept is a handheld monophonic synthesizer with internal speaker, operated using a stylus on printed circuit board pads in the pattern of a piano keyboard. The pads are terminals in a resistor ladder which forms part of a voltage controlled (relaxation) oscillator. Early devices used a programmable unjunction transistor (PUP), later superseded by a 555 oscillator and, in the current public product, a single chip.

We have produced multiple 'modded' older generation Stylophones. The stylus position can be tracked by measuring a voltage from the resistor ladder, or by pitch tracking the waveform generated by the relaxation oscillator. Control of the oscillator, or direct injection into the output stage, is also straightforward. By removing the resistor modules and replacing them with pin headers, we can plug in polyphonic capacitive touch sensing or an automated stylus. In all cases we couple with an Arduino (a 'teensy') for the algorithmic processing and MIDI i/o.

III. EVENTS

For the Tate Modern event and Oxford Ideas Festival we connected pitch-tracked Stylophones into Ableton Live and Logic Pro, with simple mathematical algorithms (e.g. Fibonacci modulo 12) triggered by note onset. Typically tapping the stylus down causes a note to sound as expected, holding it down results in the sequence. Playing the Stylophone enabled live interaction with the algorithms running behind the scenes, for example initiating multiple fragments of note sequences sounding simultaneously. This provided a starting point to discuss and demonstrate the relationships between mathematics, music and algorithms. These were popular exhibits which attracted players of all ages and musical experience.

An interesting outcome arose from using the Stylophone interface. By moving the stylus in a repeated pattern around the PCB pads, the 'drawing' pattern interacts with (or disrupts) the algorithm. It appears to be natural to draw patterns with the stylus, and this sonification of drawing patterns coupled with algorithms suggests an avenue of future work.

ACKNOWLEDGMENT

Thanks to our collaborators Emily Howard, Lasse Rempe-Gillen and Pip Willcox. The 'numbers into notes' concept for public engagement is due to Emily Howard and is based on an 'Ada sketches' performance at the Science Museum in 2015.

REFERENCES

[1] David De Roure, Pip Willcox, and Alan Chamberlain. Lovelace's Legacy: Creative Algorithmic Interventions for Live Performance. In Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion (AM'18). ACM, New York. DOI: 10.1145/3243274.3275380

[2] David McNamee. "Hey, what's that sound: Stylophone | Music". https://www.theguardian.com/music/2009/jul/06/whats-that-sound-stylophone. Retrieved 15 November 2019.

*Supported by "Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption", EPSRC award EP/L019981/1.

**DMRN+14: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2019**

***QUEEN MARY UNIVERSITY OF LONDON
TUE 17 DECEMBER 2019***

Context-Aware Audio QoE: A Case Study on the Apollo 11 Audio Archive

Alessandro Ragano[1,2,3], Emmanouil Benetos[2,3], and Andrew Hines[1]

[1]Insight Centre for Data Analytics, University College Dublin, Ireland
[2]School of EECS, Queen Mary University of London, UK. [3]The Alan Turing Institute, UK

Abstract— **The Apollo 11 mission is a great achievement of mankind. Every moment of this historic mission can be listened to through the Apollo audio archive provided by NASA. Researchers are using the Apollo archive in the field of speech processing. However, the speech quality differences among the different acoustic conditions are not explored yet. We propose a context-aware subjective quality assessment based on the differences across the mission status. The listening test goal is to collect ground truth for quality of experience (QoE) computational models aimed to assess the perceived audio quality.**

I. INTRODUCTION

The Apollo 11 mission is one of the most important of mankind's achievements and lasted 8 days 3 hours 18 minutes and 35 seconds. Beyond the historical importance, the Apollo corpus represents a fruitful resource for researchers in the field of speech and language technology. In particularly, several speech processing tasks produced unreliable results on this corpus due to the following issues: high channel noise, attenuated signal bandwidth, transmission noise, cosmic noise, analog tape noise, tape ageing noise and mostly due to the level change over time and channels of the noise conditions and the signal-to-noise-ratio (SNR). We propose to assess speech quality across the acoustic differences of the mission status by means of a listening test. Our goal is to develop a no-reference objective audio quality metric that will use the ground truth collected with the listening test and the Apollo corpus as a case study of assessing the quality of heterogeneous and historical audio archives.

II. RELATED WORK AND MOTIVATION

The Apollo corpus has been used for several tasks in speech processing and analysis, including speech activity detection (SAD) [1,2], automatic speech recognition (ASR) [2,3], speaker identification (SI) [2,4], speech to text [3], and sentiment detection [2]. No task considered the impact of the signal differences across the different mission status e.g., Earth orbit, lunar orbit, lunar surface, etc. The mission status changes the acoustic scenario due to different atmosphere, gravity, astronaut voice production [5], varying noise conditions and varying SNR levels. In addition, informal listening tests suggest a dramatic change in terms of speech quality between the Earth and the Moon. The listening test aims to assess the speech quality according to the way the context changes across each mission status. The Apollo corpus is an appropriate scenario for developing context-aware audio quality assessment models given the above-mentioned issues. By following the quality of experience (QoE) audio archive framework [5] we address quality assessment for researchers. We want to understand what are the causes that affect quality in each part of the mission. This knowledge will allow us to create a quality-score labelled dataset that covers a broad range of degradation conditions.

III. PROPOSED METHOD

By changing the degradation conditions according to the mission status, we ask the participants 4 different questions: 1) How hard is to detect speech for SAD. 2) How hard is to understand the spoken word using the absolute category rating (ACR) for ASR. 3) Given three speakers A, B, and a reference, how far are A and B from the reference for SI. This ground truth will help the development of deep learning models which can be trained jointly with the models that achieve the actual tasks. The last question is 4) how the overall quality is perceived and what are the factors that affect the overall quality. The overall quality ground truth will allow us to label the dataset for an objective quality metric and will give us the knowledge to understand the QoE among the different acoustic conditions.

REFERENCES

[1] Ziaei, A., Kaushik, L., Sangwan, A., Hansen, J.H. and Oard, D.W., 2014. Speech activity detection for NASA Apollo space missions: Challenges and solutions. In *Interspeech 2014*.

[2] Hansen, J.H., Joglekar, A., Shekhar, M.C., Kothapally, V., Yu, C., Kaushik, L. and Sangwan, A., 2019. The 2019 inaugural fearless steps challenge: A giant leap for naturalistic audio. In *Interspeech* 2019.

[3] Kaushik, L., Sangwan, A. and Hansen, J.H., 2017, August. Multi-Channel Apollo Mission Speech Transcripts Calibration. In *Interspeech*.

[4] Yu, C. and Hansen, J.H., 2017. A study of voice production characteristics of astronuat speech during Apollo 11 for speaker modeling in space. *The Journal of the Acoustical Society of America*, *141*(3), pp.1605-1614.

[5] Ragano, A., Benetos, E. and Hines, A., 2019. Adapting the Quality of Experience Framework for Audio Archive Evaluation. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE.

* This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number 17/RC-PhD/3483 and 17/RC/2289_P2 and was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1. EB is supported by RAEng Research Fellowship RF/128 and a Turing Fellowship.

The Retro in Digital: Understanding the Semantics of Audio Effects

Gary Bromham[1*], David Moffat[2], Mathieu Barthet[1] and György Fazekas[1]

[1*]Centre for Digital Music, Queen Mary University of London, UK, g.bromham@qmul.ac.uk
[2]Interdisciplinary Centre for Computer Music Research, University of Plymouth, UK

***Abstract*— It is not uncommon to hear musicians and audio engineers speak of warmth and brightness when describing analog technologies such as vintage mixing consoles, multitrack tape machines, and valve compressors. What is perhaps less common, is hearing this term used in association with retro digital technology. A question exists as to how much the low bit rate and low-grade conversion quality contribute to the overall brightness or warmth of a sound when processed with audio effects simulating early sampling technologies. These two dimensions of timbre are notoriously difficult to define and more importantly, measure. We present a subjective user study of brightness and warmth, where a series of audio examples are processed with different audio effects.**

I. Introduction

One area that dominates any discussion about retro digital technology is the lo-fi (low fidelity) aesthetics produced by early hardware samplers such as the Fairlight CMI, Akai Linn MPC 60 or Emu SP12. Our study investigates how much the low bit rate and low-grade conversion quality contributed to the sonic character of these iconic instruments. The purpose of this study is to identify the perceived impact of a range of audio effect processing techniques on the semantic terms of ‘brightness’ and ‘warmth’. A wide range of published studies were considered in our research. Researchers have considered the impact of audio effects [1], including dynamic range compression [2] and equalization [3]. Previous work on the use of semantic descriptors and their relationship to timbre perception has also been explored [4, 5].

II. Method

An online listening test was conducted where participants rate a series of audio samples on the perceived level of brightness and warmth. The samples were all processed using one of seven different audio effects, including bit reduction, dynamic range compression, equalization, and unprocessed. A further task in our study is for participants to describe why they made specific choices. This may allow us to ascertain how listeners verbalise the most prominent perceptual effects of the transformations.

III. Discussion

The results show that there is an interaction effect on both brightness and warmth between audio effect processing and instrument selection. Interestingly, 8-bit reduction tends to increase brightness but the converse is true for 12-bit processing. This could be explained by the fact that the most significant effect is seen in the case of bass sounds. Results also show that there is a significant effect of changing the audio effect processing on the dimension of brightness and warmth. The distortion style effect of the 12-bit processor is less pronounced but the effects of band limiting are noticeable enough to be perceived as warmer.

IV. Conclusion

As a sound is transformed through bit depth reduction, the inherent timbral qualities will change. There will be a resultant change in high and/or low frequency content which can be perceived as more, or less, warm or bright.

Lo-fi aesthetics produced by early digital samplers are now revered in a nostalgic way, often referred to as warm, despite once being criticised as cold and sterile. Audio effects that attempt to replicate this condition are often measured in terms of their ability to ‘transport the listener back in time’. Relative, and comparative, warmth and brightness are effective scales of evaluating such phenomena.

References

[1] David Moffat, David Ronan, and Joshua D. Reiss. 2015. An Evaluation of Audio Feature Extraction Toolboxes. In Proc. 18th International Conference on Digital Audio Effects (DAFx-15).

[2] Gary Bromham, Dave Moffat, Mathieu Barthet, and György Fazekas. 2018. The impact of compressor ballistics on the perceived style of music. In Audio Engineering Society Convention 145. Audio Engineering Society.

[3] Udo Zölzer. 2008. Digital audio signal processing. John Wiley & Sons.

[4] Andy Pearce, Tim Brookes, and Russell Mason. 2017. Timbral attributes for sound effect library searching. In Audio Engineering Society Conference: 2017 AES International Conference on Semantic Audio.

[5] Thomas Wilmering, György Fazekas, and Mark B Sandler. 2012. High-level semantic metadata for the control of multitrack adaptive digital audio effects. In Audio Engineering Society Convention 133. Audio Engineering Society.

* This research was supported through the following EPSRC project: Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption (EP/L019981/1).

**DMRN+14: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2019**

***QUEEN MARY UNIVERSITY OF LONDON
TUE 17 DECEMBER 2019***

Computational Comparison Between Different Styles of Singing Voice in Terms of the Pitch

Yukun Li[1*] and Simon Dixon[1]

[1*]Queen Mary University of London, UK, yukun.li@qmul.ac.uk

***Abstract*— This proposal presents a proposed automatic framework which aims to compare different styles of music by the quantitative method. The approach is to measure the expressive features of the pitch trajectories of the singing voice with the help of MIR. To realize this goal, this project can be separated into several processes, singing voice separation, note transcription, pitch trajectories modelling, features extraction and measurement, and music style comparison by statistic analysis.**

I. INTRODUCTION

Comparison between different styles of music has been an concerned topic in musicology. The singing voice, containing many perceived features such as pitch, rhythm, pronunciation, timbre, dynamics, etc, can be used to distinguish the music of different styles. To compare different styles in a quantitative approach, previous research[1][2] measured these features semi-automatically based on small scale recordings. With the maturity of MIR technique, extracting and measuring these features of the singing voice automatically on a large scale is promising. Thus, this proposal seeks to develop methods to realize this goal. There lists several possible approaches and potential challenges in the process.

II. PITCH TRACK EXTRACTION

Most of the music recordings are polyphonic audio, so the singing voice separation is necessary to be applied in this project to get the vocal track. Recently, there is an open-source implementation called “open-unmix”[2], which is claimed to have the state-of-art performance. This project plan to investigate this model and use it to get the monophonic vocal audio. Afterwards, PYIN, a note transcription algorithm, is used to acquire the pitch track of the monophonic audio.

III. MODELLING PITCH TRAJECTORIES

The idea of the model is to separate the pitch trajectories into several components hierarchically then do curve fitting to every part. Firstly, the model segment pitch trajectories into notes. This project aims to propose new methods to fit different styles of music. Secondly, the vibrato is detected, extracted and modeled from the pitch trajectories note by note. Thirdly, the rest part of the pitch trajectories is separated into three components that are attack, stable and release. The model fits the curve of pitch contour of them separately.

IV. FEATURES OF THE PITCH TRAJECTORIES

The features of vibrato are the rate, extent, waveform and regularity. The transition part can be downward or upward, so the slope and duration of the curve can be calculated. The stable part, approximate to a line segment, also can be characterized by the slope and duration.

V. DATASET

Currently, there are a number of collections of music which contains the singing voice. For folk music, most of data are field recording albums of many different cultures in the world. In terms of Pop music, Billboard Hot 100 charts and Million Song Dataset[4] are frequently used. For classical music, there are a number of commercial recordings of western opera.

VI. CHALLENGES

For singing voice separation part, the separated vocal track can have noises from the mixed audio, which cause errors in pitch analysis. Also, there is no ground truth to evaluate the performance of the separation automatically. To address these we can check the separation performance by listening several songs in one category of music and use large data to attenuate the influence of noises.

For modelling part, in the case of folk music, there is even no agreement in note segmentation. This project should set a rule of note segmentation to fit to the proposed model which assumes that there are the transition and stable part in one note.

The data from different source should be balanced to separate personal style from the general style.

REFERENCES

[1] Sundberg, Johan, et al. "Acoustical study of classical Peking Opera singing." Journal of Voice 26.2 (2012): 137-143.

[2] Hallqvist, Hanna, Filipa MB Lã, and Johan Sundberg. "Soul and musical theater: a comparison of two vocal styles." Journal of Voice 31.2 (2017): 229-235.

[3] Stöter, Fabian-Robert, et al. "Open-unmix-a reference implementation for music source separation." (2019).

[4] Bertin-Mahieux, Thierry, et al. "The million song dataset." (2011): 591-596.

*Research supported by CSC Foundation.