DMRN+11: Digital Music Research Network

One-day Workshop 2016

Arts One Lecture Theatre

Queen Mary University of London

Tuesday 20th December 2016

Chairs

Panos Kudumakis and Mark Sandler

DMRN+11: Digital Music Research Network One-day Workshop 2016

Queen Mary University of London

Tue 20 December 2016

Programme

10:30 **Registration** opens
Tea/Coffee

11:00 **Welcome** and opening remarks
Prof. Mark Sandler (Director, Media and Arts Technology, Queen Mary University of London)

11:10 KEYNOTE
"Rapid API - Interactive approaches to coding cross-platform digital musical instruments with machine learning", **Dr Mick Grierson (Goldsmiths University of London)**

11:50 "Towards intuitive music creation tools for musically untrained people", **Tetsuro Kitahara (Nihon University, Japan)**

12:10 "Working toward computer-augmented music traditions", **Bob L. Sturm (Queen Mary University of London) and Oded Ben-Tal (Kingston University)**

12:30 "Exploring nonlinear dynamics in musical instruments", **Tom Mudd, Simon Holland and Paul Mulholland (Open University)**

12:50 **Buffet Lunch**, Networking
Posters will be on display

14:00 "Note onset detection based on a spectral sparsity measure applied to strings instruments", **Mina Mounir and Toon van Waterschoot (KU Leuven)**

14:20 "Automatic detection of metrical structure changes", **Elio Quinton, Ken O'Hanlon, Simon Dixon and Mark Sandler (Queen Mary University of London)**

14:40 "Melody generation with stratified probabilistic context-free grammars", **Ryan Groves (Universidad del País Vasco) and Darrell Conklin (Universidad del País Vasco & Basque Foundation for Science)**

15:00 "Meter detection from music data", **Andrew McLeod and Mark Steedman (University of Edinburgh)**

15:20 **Tea/Coffee**
Posters will be on display

15:40 "Interacting with robots as performers and producers of music", **Alan Chamberlain (University of Nottingham), Kevin R Page, David De Roure, Graham Klyne and Pip Willcox (University of Oxford)**

16:00 "Classification of piano pedaling techniques using gesture data from a non-intrusive measurement system", **Beici Liang, Gyorgy Fazekas, Andrew McPherson and Mark Sandler (Queen Mary University of London)**

16:20 "Conceptualizing relevance for music information retrieval", **David M. Weigl (University of Oxford)**

16:40 "Community-building to support and encourage women and girls in music technology", **Amy V. Beeston (University of Sheffield), Lucy Cheesman (Sheffield Hallam University) and Elizabeth Dobson (University of Huddersfield)**

17:00 **Panel Discussion**

17:30 **Close***

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

DMRN+11: Digital Music Research Network One-day Workshop 2016

Queen Mary University of London

Tue 20 December 2016

Posters

1. "An empirical approach to relationship between emotion and music production quality", **David Ronan, Joshua D. Reiss (Queen Mary University of London) and Hatice Gunes (University of Cambridge)**

2. "Improved pitch trajectory estimation for polyphonic single-channel audio mixtures", **Alejandro Delgado-Castro and John E. Szymanski (University of York)**

3. "Performable spectral synthesis via low-dimensional modelling and control mapping", **William Wilkinson, Dan Stowell and Joshua D. Reiss (Queen Mary University of London)**

4. "Understanding creativity and autonomy in music performance and composition: A proposed 'toolkit' for research and design", **Alan Chamberlain (University of Nottingham), David De Roure, Pip Willcox (University of Oxford), Steve Benford and Chris Greenhalgh (University of Nottingham)**

5. "Automatic control of the dynamic range compressor using a regression model and a reference sound", **Di Sheng and György Fazekas (Queen Mary University of London)**

6. "Computational corpus analysis of native American music", **Kerstin Neubarth (Canterbury Christ Church University), Daniel Shanahan (Louisiana State University) and Darrell Conklin** (**Universidad del País Vasco & Basque Foundation for Science)**

7. "Towards a music language model for audio analysis", **Adrien Ycart and Emmanouil Benetos (Queen Mary University of London)**

8. "Designing a highly expressive algorithmic music composition system for non-programmers", **Matt Bellingham, Simon Holland and Paul Mulholland (University of Wolverhampton)**

9. "Explaining predictions of machine listening systems", **Saumitra Mishra, Bob L. Sturm and Simon Dixon (Queen Mary University of London)**

10. "Experimental digital humanities: Creative interventions in algorithmic composition on a hypothetical mechanical computer", **David De Roure, Pip Willcox (University of Oxford) and Alan Chamberlain (University of Nottingham)**

11. "Subjective evaluation of synthesised sound effects", **David Moffat and Joshua D. Reiss(Queen Mary University of London)**

12. "Comparative evaluation of rhythm transcription algorithms on polyphonic piano datasets", **Eita Nakamura and Kazuyoshi Yoshii (Kyoto University**)

13. "Intelligent audio mixing using deep learning", **Marco Martinez and Joshua D. Reiss (Queen Mary University of London)**

14. "Understanding the social character of metadata in music production", **Glenn McGarry (University of Nottingham)**

15. "Automatic transcription of vocal quartets", **Rodrigo Schramm and Emmanouil Benetos (Queen Mary University of London)**

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Towards Intuitive Music Creation Tools for Musically Untrained People

Tetsuro Kitahara[1*]

[1*]College of Humanities and Sciences, Nihon University, Japan,
kitahara@chs.nihon-u.ac.jp

***Abstract*— Main issues in achieving a system that enables musically untrained people to create music are a human interface and automatic music generation. We present two attempts concerning these issues.**

I. INTRODUCTION

Creating music is a wonderful way to enjoy music. However, this is not easy for most people because it requires expert knowledge of music. It is therefore a goal to be reached to enable musically untrained people to create their own music through computing technologies.

To reach this goal, we have two issues:
• **Human interface** Users need to be able to input their musical ideas in their mind in an easy, intuitive way.
• **Automatic music generation** Users' inputs may be too abstract or partly musically incorrect. From such inputs, the system need to generate a musical piece.

Here, we discuss these issues by presenting two of our recent studies [1], [2].

II. DRAWING-BASED MELODY CREATION

Suppose a situation that a user tried to create a melody with an automatic music composer but the generated melody is not partly satisfactory. The user has to edit the melody to obtain a satisfactory melody with music software such as a MIDI sequencer, but it is not easy. We focus on a *melodic outline*, an abstract melody representation as a manipulation object (Figure 1). Given a melody, its outline is displayed. The user can freely redraw this outline. Then, a new melody according to the outline is immediately generated.

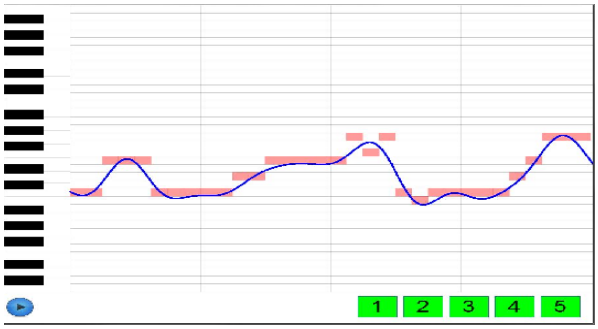


Figure 1. Screenshot of our melody creation system

*Research supported by JSPS KAKENHI Grant Numbers 26240025, 26280089, 16H01744, and 16K16180.

Tetsuro Kitahara is with College of Humanities and Sciences, Nihon University, Tokyo, Japan (e-mail: kitahara@chs.nihon-u.ac.jp)

III. SMART LOOP SEQUENCER

A loop sequencer is a useful tool especially for creating techno music because the user can create music only by choosing music loops prepared in advance. However, choosing music loops is often time-consuming because a loop sequencer has a huge number of loops. We focus on the *degree of excitement*. In techno music, the temporal evolution of excitement is an important factor and is expressed by adding/subtracting music loops. In our system, the user can directly input the temporal evolution of excitement as a curve (Figure 2). Then, the system automatically choose music loops that produce this excitement the best.

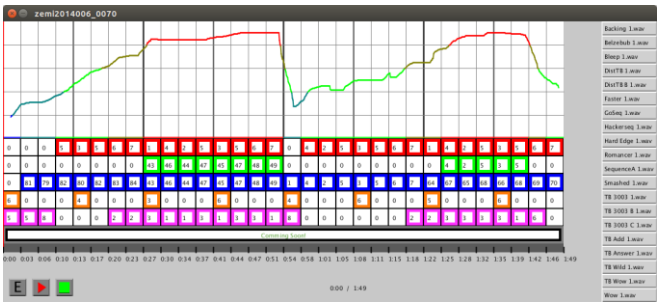


Figure 2. Screenshot of our smart loop sequencer

IV. DISCUSSION

For both systems, we adopted a temporal curve as an input object. Curves can be easily and intuitively input especially on a touch-screen computer. Inputing a curve takes very short time, so users can explore music creation by trial and error.

From the automatic music generation perspective, both systems use hidden Markov models (HMMs). User inputs are regarded as observations, and music instances (melodies, choices of music loops) to be output are regarded as hidden state sequences. Given a user input (a melodic outline, a curve of excitement), the most likely hidden state sequence is estimated. Then, the melody or audio signal corresponding to the sequence is output. In addition, it is important to consider a balance between the closeness to user inputs and musical appropriateness. The former is modeled as the emission probabilities and the latter is modeled as the transition probabilities in our HMMs.

REFERENCES

[1] Y. Tsuchiya and T. Kitahara: "Melodic Outline Extraction Method based on Non-note-level Melody Editing," Proc. SMC 2013, pp.762-767, 2013.
[2] T. Kitahara et al.: "A Loop Sequencer That Selects Music Loops based on the Degree of Excitement," Proc. SMC 2015, pp.435-438, 2015.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Working Toward Computer-Augmented Music Traditions

Bob L. Sturm[1*] and Oded Ben-Tal[2]

[1*]Centre for Digital Music, Queen Mary University of London, UK, b.sturm@qmul.ac.uk
[2]Department of Music, Kingston University, UK

***Abstract*— We discuss our work in modelling and generating music transcriptions using deep recurrent neural networks. In contrast to similar work, we focus on creating a rich evaluation methodology that seeks to address questions related to what a model has learned about the music, how useful it is for music practices, and its broader implications for music tradition. We engage with a specific homophonic music practice (session music), and present several examples of using our models for music composition in and out of the conventions of that idiom. We are currently exploring how these computer models can contribute to the tradition by engaging with its practitioners.**

The work in [1][2] describes our application of deep recurrent neural networks (specifically, long short-term memory) to modelling textual music transcriptions. Each model estimates the joint probability distribution over a set of symbols, either single characters in a large sequence of text, or higher-level tokens in single transcriptions. We build these models from crowd-sourced transcriptions of "session music," e.g., Celtic and Morris. Our code and data are available: https://github.com/IraKorshunova/folk-rnn.

Training a model entails adjusting its weights such that it predicts with high probability the "correct" symbol given a number of preceding symbols. The resulting models are generative in nature, and so can generate new transcriptions via sampling and feedback. This provides us with opportunities to discover what a model has actually learned, but also to incorporate them into a music practice "pipeline." In this work, we describe our approaches to evaluation, and how we are using the system to engage the session music community.

The typical approach to quantitatively evaluating these models involves a comparison of descriptive statistics of the training data and generated data. This provides low-level sanity checks on the model output, but does not provide any indication of the music quality of the transcriptions. We find that our models are generating transcriptions that share many of the characteristics of the real ones, but are biased to generating transcriptions with specific meters and lengths.

To move closer to evaluating the generated transcriptions, we look at them through the lens of music analysis, e.g., as a teacher of composition judging the creation of a student. We find that many generated transcriptions are plausible within the conventions of session music, e.g., many have correctly counted measures, and have a conventional AABB form with repetition and variation throughout. Many transcriptions suffer from poor harmonic implications, however, and lack coherence between sections.

To test the general abilities of each model, we examine the response of the system to seed material that is musically valid, but outside the conventions of the training data ("nefarious testing"). For instance, we test whether the model can actually count time, or has learned to repeat and vary material. We see that the model can do these things, but only within a narrow context. Its knowledge lacks generality.

To test how applicable these models are to composition, we use them in an iterative process of generation and curation, both in and out of the conventions of the training data. We have produced and posted online many examples of this, e.g., https://soundcloud.com/sturmen-1. We find particular interest in the "failures" of the models, which calls into question the compatibility of the training the systems to faithfully reproduce real transcriptions.

We are expanding upon this work in several ways. We have generated many thousand transcriptions without curation and invited session music practitioners to play and comment, e.g., https://highnoongmt.wordpress.com/2016/09/12/folk-rnn-session-tunes-volume-1-of-10. This work is on-going but interesting discussions have already occurred, e.g., https://thesession.org/discussions/39604. We are exploring different ways of sampling from the model output, and ways of involving a critic system in the model training. We also intend to train the models starting from their current weights on more specific subsets of the session practice, e.g., Irish jigs, or of other music traditions, e.g., gospel music.

ACKNOWLEDGMENT

João Felipe Santos and Iryna Korshunova.

REFERENCES

[1] B. L. Sturm, J. F. Santos, O. Ben-Tal, and I. Korshunova, "Music transcription modelling and composition using deep learning," in *Proc. 1st Conf. Computer Simulation of Musical Creativity*, 2016.

[2] B. L. Sturm and O. Ben-Tal, "Evaluating music transcription modelling and composition with a token-based deep recurrent neural network," *J. Creative Music Systems* (in review) 2016.

*This research is supported by AHRC Grant No. AH/N504531/1.

BLS is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, UK (corresponding author e-mail: b.sturm@qmul.ac.uk).

OBT is with the Music Department, Kingston University, UK (e-mail: O.Ben-Tal@kingston.ac.uk).

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Exploring nonlinear dynamics in musical instruments

Tom Mudd[1*], Simon Holland[1], Paul Mulholland[2]

[1*]Music Computing Lab, The Open University, UK, tom.mudd@open.ac.uk
[2]Knowledge Media Institute, The Open University, UK

***Abstract*— This paper examines a particular mechanism in musical interactions: nonlinear dynamical processes. The subject is related to ideas of surprise and exploration in interactions with creative tools more broadly.**

I. BACKGROUND

Nonlinear dynamical processes are at the heart of many acoustic musical instruments, play a part in many electronic musical instruments, and have been deliberately implemented in a range of digital musical instruments. Interactions with such processes are the rule rather than the exception in our everyday interactions with the physical world, yet these processes have rarely been explicitly examined from an interaction perspective.

In the context of musical instruments, the term 'dynamical' refers to an instrument's state being dependent on not only the current inputs, but also on previous inputs. `Nonlinear' implies that the function relating previous states to the current state includes a nonlinear term. Where a system is both nonlinear and dynamical, it may exhibit a wide range of behaviours in response to changes in input: abrupt transitions, instability, locking into particular states, producing very diverse outputs for the slightest change at the inputs. Interactions with nonlinear dynamical processes can therefore be potentially confusing, unstable and unpredictable. However, as acoustic instruments demonstrate, this is not to say that they are not controllable, and are indeed sites where considerable skill can be developed over long time periods.

II. INTERACTION

In human-computer interaction (HCI) there is a tendency to make input-output relationships as clear and as simple as possible. This approach carries over into the design of digital tools for music and other kinds of creative activity, resulting in a markedly different kind of creative engagement. However, the complexity of interaction with acoustic instruments may be linked to the long term potential for exploration in such instruments. Magnusson [1] makes this connection, pointing out that these complexities are inherent in the physical materials: ``the physics of wood, strings and vibrating membranes were there to be *explored* and not invented.", emphasis in original). Magnusson explicitly relates this to the nonlinear behaviour of such instruments:

"As opposed to the generic explicitness of the digital instrument, the acoustic instrument contains a boundless scope for exploration as its material character contains a myriad ways for instrumental entropy, or `chaotic', non-linear behaviour that cannot be mapped and often differs even in the same type (brand and model) of instruments."

In a similar vein, Strange and Strange [4] highlight bowed interactions as an aspect of violin playing which is still being explored by performers and composers, and is yet to be exhausted. The bowed interaction is a key source of nonlinearity in violin physics [3].

III. STUDIES

User studies were conducted with a set of interfaces designed to test the influence of the presence of nonlinear dynamical processes [2]. Two interfaces were designed and implemented that included these processes, alongside two further interfaces that were designed to be similar but not include such processes. The nonlinear dynamical interfaces were seen as both more surprising ($p < 0.001$), and offering more scope for exploration and discovery ($p < 0.001$). Furthermore, and perhaps more surprisingly, they were not seen as less controllable than the other interfaces.

A subsequent ethnographic study approached the issue from a different angle. Interviews were conducted with a range of musicians engaged in contemporary practices that embrace surprise, unpredictability, and a two-way dialog with their musical tools. These interviews examined how these musicians use their tools, how they think about interaction, their views on the relationships between surprise, control and exploration, and the role that nonlinear dynamics may play in these interactions.

REFERENCES

[1] T. Magnusson (2009). Of Epistemic Tools: musical instruments as cognitive extensions. *Organised Sound*, Vol. 14 (2), Cambridge University Press, 168–176.
[2] Mudd, T., Holland, S. and Mulholland, P. (2015). Investigating the effects of introducing nonlinear dynamical processes into digital musical interfaces. In: *Proceedings of Sound and Music Computing Conference*, Maynooth, Ireland.
[3] J. O. Smith (2010). Physical Audio Signal Processing. Online book, accessed 10 January 2016, available at http://ccrma.stanford.edu/~jos/pasp/
[4] P. Strange and A. Strange (2001). The Contemporary Violin: Extended Performance Technique. University of California Press.

*All authors are based at The Open University, Milton Keynes, MK7 6AA (corresponding author e-mail: tom.mudd@open.ac.uk).

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Note Onset Detection based on a Spectral Sparsity measure applied to Strings Instruments

Mina Mounir[1*] and Toon van Waterschoot[1*]
[1*]Department of Electrical Engineering (ESAT), KUleuven, Belgium,
{mina.mounir, toonvanwaterschoot}@esat.kuleuven.be

***Abstract*— Being the atomic component for a melody's time dimension, the detection of note onsets is gaining a growing interest in the arising research fields music information retrieval (MIR), machine listening and music processing. We propose a new note onset detection algorithm NINOS^2 exploiting the spectral sparsity difference between different parts of a musical note. Added to the outstanding performance of NINOS^2 when applied to automatically annotated guitar melodies and chords progression [1], the proposed algorithm consistently outperforms the state-of-the-art LogSpecFlux (LSF) for the sustained-strings group of instruments crossing the 50% F1-score border. We also propose an additional performance measure to assess the relative position of detected onsets w.r.t. each other.**

I. METHOD SUMMARY

In this article, we present NINOS^2, an algorithm that proposes a new pre-processing and reduction function steps for calculating a smooth Onset Detection Function (ODF) – a highly subsampled version of the original music signal having distinguishable amplitude peaks at time instants where onsets appear [2] – making it easier for the peak picking step to identify the onsets.

For the pre-processing, the subset of low-energy frequency coefficients – traditionally filtered out as they were considered as noise to the detection process – is selected for the ODF calculation. Moreover, NINOS^2 exploits the difference in spectral sparsity between the transient and the following steady-state component of a musical note when calculating the ODF noted ℵ(i):

$$\aleph(i) = \frac{1}{\sqrt[4]{J}} \frac{\|Y_i\|_2^2}{\|Y_i\|_4}. \qquad (1)$$

where Y is the pre-processed frequency magnitude spectrum, i is the frame index and J is the count of selected coefficients. It is an energy and inverse-sparsity measure as onsets are usually accompanied by an energy increase and they mark the start of transients which are spectrally less sparse than the steady-state part of a note.

[*]This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven Research Council CoE PFV/10/002 (OPTEC), KU Leuven Impulse Fund IMP/14/037, the FP7-PEOPLE Marie Curie Initial Training Network "Dereverberation and Reverberation of Audio, Music, and Speech (DREAMS)" funded by the European Commission under Grant Agreement no. 316969, IWT O&O Project nr. 150432 "Advances in Auditory Implants: Signal Processing and Clinical Aspects", and IWT O&O Project nr. 150611 "Proof-of-concept of a Rationed Architecture for Vehicle Entertainment and NVH Next-generation Acoustics (RAVENNA)". The scientific responsibility is assumed by its authors.

II. EXPERIMENTAL RESULTS

We ran the experiment on a variety of strings instruments: guitar, violin, viola, cello, …etc. The dataset is generated and automatically annotated as described in [1]. Table I and Fig.1 show a snippet from the result. In Fig.1, onsets ground-truth are marked with vertical squares, true positives are marked with circles while false positives with x marks.

TABLE I. PERFORMANCE COMPARISON (BEST-F1 SCORE)

Instruments group	*NINOS^2*	*LSF*
Plucked Strings	0.9861	0.9828
Sustained Strings	0.5724	0.4878

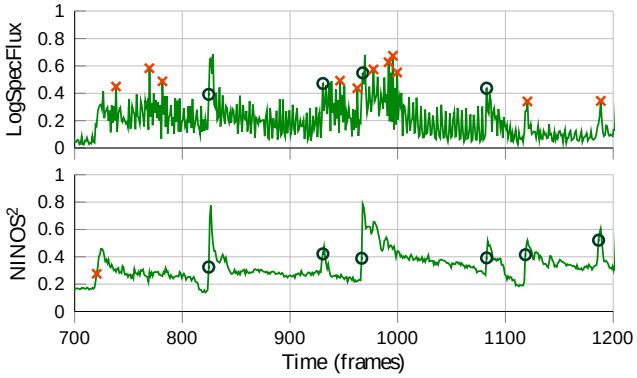


Figure 1. Comparison of ODF's and Detections for a *Cello*.

REFERENCES

[1] M. Mounir, P. Karsmakers, and T. van Waterschoot, "Guitar note onset detection based on spectral sparsity measure," in *Signal Processing Conference(EUSIPCO),2016Proceedingsofthe24thEuropean*. IEEE, 2016, pp. 978–982

[2] P. Leveau and L. Daudet, "Methodology and tools for the evaluation of automatic onset detection algorithms in music," in *Proc. Int. Symp. Music Information Retrieval*, 2004, pp. 72–75.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Automatic Detection of Metrical Structure Changes

Elio Quinton[1*], Ken O'Hanlon[1*], Simon Dixon[1*], Mark Sandler[1*]

[1]Queen Mary University of London, UK, e.quinton@qmul.ac.uk

***Abstract*— Meter inference algorithms are typically designed to track metrical structure in presence of mild deviations of the feature estimates over time in order to account for performance imprecisions, expressive timing or musical effects such as accelerando. Abrupt changes of metrical structure over time are comparatively rarely addressed. In this paper, we present an unsupervised approach to detect metrical structure changes. Formulating the problem as a metrical structure based segmentation retrieval task, we present a variant of sparse Non-negative Matrix Factorisation (NMF) and demonstrate state of the art performance.**

I. INTRODUCTION

Meter inference algorithms are typically designed to track metrical structure in presence of mild deviations of the feature estimates over time in order to account for expressive timing or performance imprecisions. Abrupt changes of metrical structure over time are comparatively rarely addressed and even less so as a task in itself. See for instance the bar pointer model [1], which allows the tracking of abrupt meter changes but requires supervised learning of the metrical structures. In this paper we propose an unsupervised method for the detection of metric modulations within a music piece. We formulate the task as a metrical structure based segmentation retrieval problem. We restrict this study to abrupt modulations from one section of stable metrical structure to another section with a different but still stable metrical structure. Segment boundaries represent the time points at which the metric modulation happens. The number, length and metrical structure of each segment is a priori unknown.

II. METHOD

We first compute two rhythmograms based on the Fourier transform and Auto-Correlation Function (ACF) of the windowed onset detection function (previously computed using the *superflux* method [2]) respectively, using 12s Hann windows with 0.24s overlap. The ACF rhythmogram is mapped to the frequency domain and we compute the *metergram* (**V**) as the element-wise product of the two rhythmograms. The distribution of energy in a metergram relates to the metrical structure of the music. Changes in metrical structure over time therefore result in apparent structure in the metergram, which we seek to recover. Segments of consistent metrical structure are expected to correspond to homogeneous regions in the metergram so that segmentation may be retrieved by detection of *homogeneity*.

We propose a variant of sparse Non-negative Matrix Factorisation (NMF) seeking to minimize the cost function:

$$D_\beta(\mathbf{WH}|\mathbf{V}) + \alpha\left(\frac{1}{\beta}\sum_{n=1}^{N}\|\mathbf{y}_n\|_\beta^\beta\right) \quad (1)$$

where **W** and **H** are the template and activation matrices respectively and $y_{k,n} = h_{k,n} \times \|\mathbf{w}_k\|_2$ [3]. We set $\beta = \frac{1}{2}$ in our experiment, which enforces a stronger sparsity constraint than a L_1 penalty. The segmentation is retrieved from the application of a Hidden Markov Model (HMM) on **H**. We compare this method (L_β-S-β-NMF) to existing NMF methods as well as a popular novelty-based segmentation approach, which we refer to as SSM-Foote [4].

III. RESULTS

The results show that homogeneity-based NMF-powered methods outperform the standard novelty-based approach. The comparatively low hit-rate F-measure obtained with the SSM Foote method suggests that NMF-based methods produce segment boundary estimates with more precise location than the novelty-based approach. In addition, it appears that L_β-S-β-NMF and ARD achieve the best performance. This suggests that very strong sparsity constraints are instrumental in successfully estimating segmentation.

TABLE I. SEGMENTATION PERFORMANCE RESULTS

Methods	***ppr***	***prr***	***pfm***	S_o	S_u	S_f	***Fm***
L_β-S-β-NMF	0.77	0.78	0.73	0.72	0.84	0.75	0.41
ARD β=1	0.70	0.91	0.75	0.69	0.70	0.70	0.42
SNMF-S	0.84	0.50	0.57	0.58	0.85	0.69	0.31
SSM Foote	0.66	0.81	0.68	0.68	0.68	0.68	0.07

REFERENCES

[1] N. Whiteley, A. Cemgil and S. Godsill, "Bayesian Modelling of Temporal Structure in Musical Audio", In Proceedings of *the International Society of Music Information Retrieval (ISMIR)*, 2006.

[2] S.Böck and G. Widmer, "Maximum Filter Vibrato Suppression for Onset Detection" in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, 2013

[3] E. Quinton, K. O'Hanlon, S. Dixon and M. Sandler, "Tracking Metrical Structure Changes with Sparse-NMF", in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 2017.

[4] J. Foote, "Automatic Audio Segmentation Using a Measure of Audio Novelty", in *IEEE International Conference on Multimedia and Expo (ICME)*, 2000

*Research supported by the EPSRC award 1325200 and the AHRC grant AH/L006820/1.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Melody Generation with Stratified Probabilistic Context-Free Grammars

Ryan Groves[1] and Darrell Conklin[1,2]

[1]Department of Computer Science and Artificial Intelligence
University of the Basque Country UPV/EHU, San Sebastián, Spain
[2]IKERBASQUE: Basque Foundation for Science, Bilbao, Spain

***Abstract*— This work builds on previous methods for melody generation using a PCFG, specifically by using a template melody, and generating a new melodic surface with a stratified grammar.**

I. INTRODUCTION

The connection between music and language has long been investigated through both analytical and computational means. Both fields deal with temporal sequences which have an inherent hierarchical structure, and contain connections between events that span a large temporal range. For this reason, linguistic techniques have been explored for their application to musical sequences. The Generative Theory of Tonal Music (GTTM) is a formative theory that used a generative grammar to notate musical structure [5]. Computational methods in the form of formal grammars have been used to both analyze music as well as generate new musical excerpts [6]. Using a grammar in a probabilistic way makes it straightforward to find the most probable parse tree [2, 1] for a given melody and also to generate new sequences [7]. Through his work, Hamanaka also created a set of tree-based music analyses using the theory of GTTM [4].

II. MODEL

This work is based on previous work on PCFGs for melodic generation [3], and presents a method for automatically generating new melodic sequences from templates of existing melodies. As the background levels in a parse tree contain more jumps in pitch, a stratified grammar was used to regenerate only the musical surface of a tree structure (based on the separately-trained surface grammar), while still maintaining a similar overall hierarchical structure to the original template tree.

III. PROCESS

We started with the PCFG model from [3], trained on a treebank of time-span tree reductions provided in the GTTM dataset. Using the same conversion algorithm as in [3], we created a bank of parse trees with a particular linked viewpoint as the representation of each node. The chosen viewpoint links key-relative pitch class interval with a metric interval. Furthermore, each tree in the training treebank was separated based on node depth. Those nodes within a close range to any leaf node were separated, and a new PCFG was created that represented only the surface rules and the distributions of their rule expansions. With a trained PCFG, it is then possible to create new trees by iteratively sampling from the rule distributions, and continually expanding nodes until the leaf nodes are reached. Because the metric interval is based on the metric levels presented in GTTM, it does not actually specify particular onset times for each note. Therefore, the actual onset times of the original melody are used as a template as well.

IV. RESULTS & DISCUSSION

The results provide a more realistic melody than the previous method of generating the entire tree. Because the background structure of a template melody is used, certain notes from the original melody will always be part of any generated melody, thus giving the generated melody a little more form. Based on a listening session of 20 examples, it seems that the new generation method generates pleasing melodic figures more often than the previous PCFG method.

REFERENCES

[1] É. Gilbert and D. Conklin. A probabilistic context-free grammar for melodic reduction. In *Proceedings for the International Workshop on Artificial Intelligence and Music, International Joint Conference on Artificial Intelligence*, pages 83–94, Hyderabad, India, 2007.

[2] R. Groves. Automatic melodic reduction using a supervised probabilistic context-free grammar. In *Proceedings of the International Conference of Music Information Retrieval*, pages 775–781, New York, NY, 2016.

[3] R. Groves. Towards the generation of melodic structure. In *Proceedings of the Fourth International Workshop on Musical Metacreation*, pages 1–8, Paris, France, 2016.

[4] M. Hamanaka, K. Hirata, and S. Tojo. Implementing "A generative theory of tonal music". *Journal of New Music Research*, 35(4):249–77, 2007.

[5] F. Lerdahl and R. Jackendoff. *A generative theory of tonal music*. The MIT Press, Cambridge, MA, 1983.

[6] F. Lerdahl and Y. Potard. *La composition assistée par ordinateur*. Rapports de recherche. Institut de Recherche et Coordination Acoustique/Musique, Centre Georges Pompidou, Paris, France, 1986.

[7] D. Quick and P. Hudak. Grammar-based automated music composition in haskell. In *Proceedings of the first ACM SIGPLAN workshop on Functional art, music, modeling & design*, pages 59–70, New York, NY, 2013.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Meter Detection From Music Data

Andrew McLeod[1] and Mark Steedman[1]

[1]School of Informatics, University of Edinburgh, UK, A.McLeod-5@sms.ed.ac.uk

***Abstract*— Meter detection is the organisation of the beats of a given musical performance into a repeating structure of trees at the bar level, in which each node represents a single note value. The metrical structure must be properly aligned in phase with the underlying musical performance so that the root of each tree properly aligns with a bar. In this talk, I will introduce a lexicalized probabilistic context free grammar designed for the task, and show that it performs well on a variety of symbolic corpora when compared to existing work. I will also discuss further enhancements that could be made to the model in order to allow it to achieve good results on a wide variety of genres, and allow it to be run on audio data. Finally, I will discuss this work's role in a larger project stressing the importance of symbolic music analysis for future work on Automatic Music Transcription.**

I. LEXICALIZED PROBABILISTIC CONTEXT FREE GRAMMAR

In designing the grammar, we were careful to make as few assumptions as possible so it can be applied to different genres of music directly (assuming that training data is available). It is a standard probabilistic context free grammar (PCFG) where each terminal is also assigned a head based on some property (in this work, note length is used). Strong heads (in this work, those representing longer notes) propagate upwards through the metrical tree. This lexicalization allows the grammar to model rhythmic dependencies rather than assuming independence as in a standard PCFG, and the pattern of strong and weak beats and sub-beats is used to determine the correct metrical structure.

II. EVALUATION

In evaluation, for each of the three levels (sub-beat, beat, and bar) of the guessed metrical tree, if it matches exactly a level of the correct metrical tree, that is counted as a true positive. On the other hand, a clash is counted as a false positive. Additionally, each of the correct metrical tree's levels which were not matched count as a false negative. Precision, recall, and F1 are then computed based on the given true positive, false positive, and false negative values.

For comparison, we created two baseline models: one which always guesses 4/4 time with the first bar starting on the first note, and a simple PCFG with no lexicalization. We also compare against the models proposed by Temperley [3] and De Haas & Volk [1], though the latter is still in progress due to technical difficulties.

In Table 1, we report the results on two corpora: 1) the 48 fugues from the Well-Tempered Clavier, and 2) the 15 Bach Inventions. We used leave-one-out cross-validation within each corpus for training, and the F1 scores reported are averaged throughout each corpus. Evaluation on additional corpora is in progress.

TABLE I. PRELIMINARY RESULTS

Method	***WTC F1***	***Invention F1***
Temperley [3]	0.63	0.58
4/4	0.45	0.58
PCFG	0.64	0.61
LPCFG	**0.83**	**0.65**

III. CONCLUSION

The proposed grammar shows promise in time signature detection, and it is clear that lexicalization helps. Future work will investigate the use of different lexical heads; for example, whether using note pitch in the heads would improve performance. We will also adapt the gram- mar to run on audio data. There are a couple of different options for how to do this: for one, an off-the-shelf onset detection model could be used to convert the audio into a symbolic format (since we do not require note pitch); alternatively, the lexicalized heads could be adapted to draw their strengths directly from the audio data based on some feature extraction.

This work is part of a larger project on using symbolic music analysis to improve automatic music transcription (AMT), and I will briefly discuss how this grammar, modeled jointly with a previously designed voice separation model [2] and other future components, might be adapted to be used alongside existing AMT models, as well as models for other music information retrieval tasks.

REFERENCES

[1] De Haas, Bas W., & Volk, A. (2016). Meter Detection in Symbolic Music Using Inner Metric Analysis. In *Proceedings of the 17th International Society for Music Information Retrieval Conference*, pp. 441–447.

[2] McLeod, A., & Steedman, M. (2016). HMM-Based Voice Separation of MIDI Performance. *Journal of New Music Research, 45*(1), pp. 17-26.

[3] Temperley, D. (2009). A Unified Probabilistic Model for Polyphonic Music Analysis. *Journal of New Music Research, 38*(1), pp. 3–18.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Interacting with Robots as Performers and Producers of Music

Alan Chamberlain[1*], Kevin R Page[2], David De Roure[2], Graham Klyne[2], Pip Willcox[2]

[1*]Department of Computer Science, MRL, University of Nottingham, UK, azc@cs.nott.ac.uk
[2]Oxford e-Research Centre, University of Oxford, UK

***Abstract*— This paper discusses some of the issues relating to Human Robot Interaction and the use of robotics in performance and music creation**

I. INTRODUCTION

Is it really so strange to think about a robot as something, or perhaps someone that can produce music, as a performer or even as a composer? What happens when robots perform on stage to live audiences, and when they are perceived as intelligent? In this abstract we start to unpack and explicate some of the issues that emerge when the fields of music technology and robotics come together. The aim of this piece of writing is to prompt the *Digital Music Research* community to engage in debate, in order develop this emerging field of research.

II. WHERE IS THE HUMAN, THE ROBOT, THE "INSTRUMENT"?

Understanding *Human Robot Interaction* (HRI) is complex, and by its very nature research in this area is multifaceted. Additional levels of complexity become part of this project, when systems are looked at that are in some way intelligent, creative and in particular when the physical form of the robot is humanoid, or the actions/tasks that the robotics are physically able to accomplish, somehow mirror human activity, or give the appearance of intelligence/sentience. Understanding where agency lies, who has ownership of the occurring actions and how control is mediated across a given system are all key to the understanding of robotic systems as musical. It is not the case that all robotic musical systems are the same. Godfried-Willem Raes takes an approach that places the human at the centre of the system, the performance and the composition. Instruments are augmented robotically and are controlled by the movements of a performer. The instruments are played (mechanically) in a way that that a human could never emulate. The robots are not necessarily mimicking human capacity or physicality. It is through the performers' 'accountable' actions that the augmented-instruments are played, and in this respect the system becomes an instrument. Understanding this relationship between human and machine as one of symbiosis, and one that affords mechanisms for music creation and performance is one that should be highlighted and needs further development in respect to developing understandings of the ways that tools are used for creative purposes in this continuum. In our next section we start to unpack the role of robots in music performance.

II. PERFORMING ROBOTS?

Can robots perform, or is it the case that humans program computers to give the impression that the robotic system is the performer? Research laboratories such as the Center for Music Technology (Georgia Tech) [2] have offered a range of robotic systems, from systems that are able to 'jam' and improvise to robotic prosthetic limbs (for drumming), but are these really able to do the things that performers do on stage, or are they akin to audio automaton? Perhaps an initial understanding of robotic musical performances could be brought about by examining the interaction between the 'performer' and audience, and by looking at different settings, dynamics and situations. It might be that audience expectations are different for someone who plays with a robotic prosthetic limb, as compared to other systems where the agency of the system is less obvious, and the audience is unsure of where this lies. Of course this is to presume that the technology used is not autonomous and has no creative agency in its own right.

III. CONCLUSION

The role that robots can play in music creation and performance, and our interaction with them, is something that is arguably still not fully understood. Robots and the application of robotics in the field of digital music performance and creation is constantly evolving, and as the technologies evolve it becomes apparent that we too need to reassess our relationship with such technologies, both in terms of application and theory.

ACKNOWLEDGMENT

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

REFERENCES

[1] Godfried-Willem Raes (November 2016) For examples of the work of this work *see*: http://www.logosfoundation.org/index-god.html
[2] Center for Music Technology (Georgia Tech) (November 2016) For examples of the work of this lab *see*: http://www.gtcmt.gatech.edu

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Classification of Piano Pedaling Techniques Using Gesture Data from a Non-Intrusive Measurement System

Beici Liang[1*], György Fazekas[1], Andrew McPherson[1] and Mark Sandler[1]

[1]C4DM, Queen Mary University of London, UK, beici.liang@qmul.ac.uk

***Abstract*— This paper presents the results of a study of piano pedaling techniques on the sustain pedal using a newly designed measurement system. This system is comprised of an optical sensor mounted in the pedal bearing block and the Bela platform for recording audio and sensor data. Using the gesture data collected from the system, the task of classifying these data by pedaling technique was undertaken using a Support Vector Machine (SVM).**

I. INTRODUCTION

The role of pedaling in piano performance is important as "the soul of the piano", according to the Russian pianist Anton Rubinstein. Pianists can add variations to the tones with the help of three pedals. However, the role of pedaling as an instrumental gesture to convey different timbral nuances has not been adequately and quantitatively explored, despite the fact that acoustic effect of the sustain pedal on piano sound has been studied [1].

Since the pedaling parameters are difficult to estimate from the audio signal, the data for this study was captured using a new measurement system. This system enables recording the pedaling gesture of real players and the piano sound under normal playing conditions. Four pedaling techniques (quarter, half, three-quarters and full pedal) were classified using continuous pedal position changes in one dimension as inputs to an SVM algorithm.

II. PEDALING TECHNIQUE STUDY

A schematic overview of the setup developed for this study is shown in Fig.1. Near-field optical reflectance sensing was used to measure the position of the sustain pedal. An Omron EESY1200 sensor was mounted in the pedal bearing block. The output voltage is proportional to light and roughly follows the inverse square of the pedal-sensor distance. After calibrating the output voltage through a customized Printed Circuit Board (PCB), the sensor data was recorded at 22.05kHz sample rate using the analog input of the Bela platform [2]. The piano sound was synchronously recorded at 44.1kHz as audio input to Bela as well. These two types of signal data were stored as CSV files and binary files respectively. Pedaling in ten excerpts of Chopin's piano music were recorded and labeled to provide a basic ground truth dataset.

*Research supported by China Scholarship Council (CSC) and Media & Arts Technology Centre for Doctoral Training (MAT CDT).

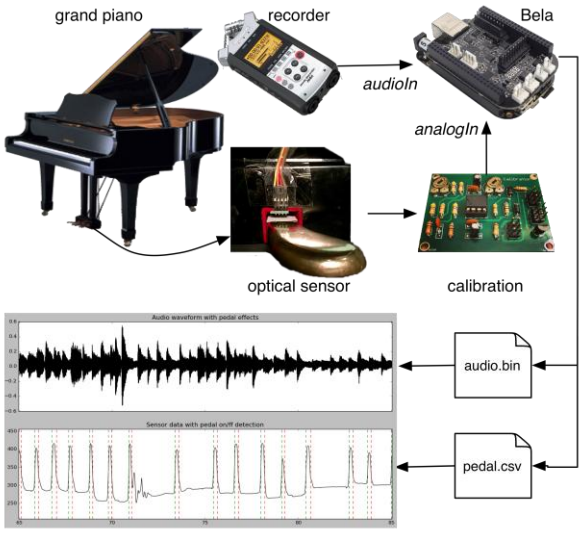


Figure 1. Scheme of the setup used for pedaling technique study.

Classification of pedaling techniques was implemented using a SVM. A Radical Basis Function (RBF) kernel SVM was trained from the raw sensor data. A mean F-measure score of 0.74 was obtained using cross validation. It can be improved to 0.93 by training a linear kernel SVM from the Gaussian features of the sensor data. In our study, the onset and offset time of each pedaling technique can also be automatically detected by setting a threshold derived from a peak detection algorithm. These results can be visualized in an audio based score following application to show pedaling together with the player's position in the score.

ACKNOWLEDGMENT

Thanks to Siying Wang for her score following demo.

REFERENCES

[1] H. M. Lehtonen, et al. "Analysis and modeling of piano sustain-pedal effects." *The Journal of the Acoustical Society of America*, 122(3), 2007, pp. 1787-1797.

[2] G. Moro, S. A. Bin, R. H. Jack, C. Heinrichs and A. McPherson. "Making high-performance embedded instruments with Bela and Pure Data." *Proc. International Conference on Live Interfaces*, Brighton, UK, 2016.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Conceptualizing Relevance for Music Information Retrieval

David M. Weigl[1*]

[1*]School of Information Studies, McGill University, Canada & University of Oxford e-Research Centre

***Abstract*— Although there have been repeated calls in the MIR literature for a greater emphasis on the (potential) users of music information systems—complementing valuable and thriving research on MIR algorithms—the impact of formal investigations of user information needs and information behaviour on MIR system design has been limited. Challenges of generalization, a lack of systematic synthesis of results, and the disconnect between system/evaluation task designers and user studies researchers have been proposed as potential reasons for this situation [1]. To address these issues, we present an analytical interface and query mechanism operating over a large and extensible corpus of empirical findings identified during a systematic analysis of 159 research articles presenting MIR user studies, coded according to a conceptual framework for relevance, a central notion at the heart of information searching and information retrieval [2].**

I. MOTIVATION

The definition and operationalisation of experiential similarity and relevance measures—a key research priority for Music Information Retrieval (MIR) [3]—stands to benefit greatly from user-focussed research efforts. Discussions on the nature of relevance and its place at the heart of Information Retrieval (IR) have been ongoing for decades, and have been traced to the very beginnings of organized academic research into textual IR. While some initial investigations have probed this notion in the music information domain, finding significant overlap with textual IR in terms of the criteria users apply in their relevance judgements [4-5], formal consideration of relevance in MIR has remained scarce. To address this, we have adapted an established conceptual framework for relevance in textual IR (the stratified model of relevance interactions [2]) to clarify the variety of relevance considerations for MIR [6].

II. CORPUS AND CODING ACTIVITY

We have conducted a systematic analysis of a set of 159 articles reporting on user studies in the MIR literature, compiled in a previous bibliometric investigation by Lee and Cunningham [1]. We investigate the content, codifying empirical findings reported within each article according to our conceptual framework. The resulting corpus currently contains 866 discrete findings derived from 176 studies. We have developed tools to enable real-time updates as new studies and findings are added to the corpus, while supporting inter-coder consistency in the choice of descriptors. The corpus can be accessed using two analytical interfaces. Respectively, they support user researchers in literature synthesis and hypothesis generation via an interactive conceptual-category co-occurrence view; and enable system designers to gain a rapid overview of MIR user research findings pertaining to their specific area of interest by allowing the corpus to be filtered using a descriptor selection interface.

III. PROMOTING THE IMPACT OF MIR USER RESEARCH

Lee and Cunningham outlined a number of challenges limiting the impact of MIR user research. These include: generalization, whereby the common reliance on convenience sampling and limitations of sample size require triangulation between different studies in order to identify findings that generalize across multiple groups of users; a lack of systematic synthesis of research results, complicated by the highly diffuse nature of their dissemination; and, perhaps as a consequence, a disconnect between system/evaluation task designers and user studies researchers. Lee and Cunningham's list of articles is publicly shared in the spirit of collaborative and transparent research synthesis. We gratefully adopt this collection, and reciprocate by making our tools and corpus of coded studies and findings available (*http://relevance.linkedmusic.org*). We invite MIR user researchers to help grow the corpus, and hope that this process will help to promote the impact of MIR user studies, and ultimately to more useful MIR systems.

*Coding activity performed in collaboration with C. Guastavino, D. Steele, and J. Bartlett at the School of Information Studies, McGill University. This work forms part of David M. Weigl's doctoral research completed at McGill. Support in part by a McGill University William Dawson Award to C. Guastavino, and by the EPSRC FAST IMPACt project (EP/L019981/1). David M. Weigl is now with the University of Oxford e-Research Centre (e-mail: david.weigl@oerc.ox.ac.uk).

REFERENCES

[1] J.H. Lee and S.J. Cunningham. The impact (or non-impact) of user studies in music information retrieval. In *Proc. ISMIR 2012*, pages 391–396, 2012.

[2] T. Saracevic. "Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II". *Journal of the American Society for Information Science and Technology*, 58(13):1915–1933, 2007.

[3] J. S. Downie. "Music information retrieval". *Annual review of information science and technology*, 37(1):295–340, 2003.

[4] C. Inskip, A. MacFarlane, and P. Rafferty. "Creative professional users' musical relevance criteria". *Journal of Information Science*, 36(4):517–529, 2010.

[5] A. Laplante. "User's relevance criteria in music retrieval in everyday life: An exploratory study". In *Proc. ISMIR 2011*, p.601-606, 2010.

[6] D. M. Weigl and C Guastavino. "Applying the stratified model of relevance interactions to music information retrieval". In *Proceedings of the Association for Information Science and Technology*, 50(1): p.1–4, 2013.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Community-building to support and encourage women and girls in music technology

Amy V. Beeston[1*], Lucy Cheesman[2] and Elizabeth D. Dobson[3]

[1*]Department of Computer Science, University of Sheffield, UK, a.beeston@sheffield.ac.uk
[2]Faculty of Arts, Computing, Engineering & Sciences, Sheffield Hallam University, UK
[3]School of Music, Humanities & Media, University of Huddersfield, UK

***Abstract* — This paper documents a community-building project underway in the north of England that aims to support and encourage women and girls in academic and industrial careers relating to music technology.**

I. BACKGROUND

The issue of gender imbalance in music technology has been receiving a lot of attention in recent years. Significantly fewer girls apply as undergraduates to study music technology (Born and Devine, 2015) or follow careers in related industries (see e.g. female:pressure's 2013 and 2015 reports). The same pattern is borne out locally and globally for established academics too. Female-led contributions to the UK-based Digital Music Research Network meetings averaged 12.3% of all contributions in the period 2011 to 2015, and occurred at a similar rate to those reported recently at conferences of The International Society of Music Information Retrieval, where 14.1% of all contributions in the period 2000 to 2015 were female-led papers (Hu et al., 2016). While girls have often been reported to lack confidence in engaging with technology, a recent study at Georgia Tech reported that integrating music to the syllabus provided a way to increase engagement in computer programming amongst girls (Freeman et al., 2014). Thus it might be the case that music offers a specific pathway along which girls and women can, in the right setting, begin to engage more widely with STEM subjects.

II. COMMUNITY-BUILDING

This paper reports on a social initiative in the north of England that engages women and girls in sound and music technology. Established in 2015, the Yorkshire Sound Women Network provides opportunities for women and girls who may have felt excluded or uncomfortable in male-dominated environments to meet, share knowledge formally and informally, and thereby develop their technical and creative skills. Some of this activity has taken place in Sheffield as part of the city's 'Year of Making' celebrations. A 'Catalyst: Festival of Creativity' grant recently funded 8 expert-led workshops, 3 peer-learning maker-space days and a monthly social gathering. Averaging 10 registrations per workshop, these sessions covered a wide range of topics: sound synthesis, machine listening, performance hardware, electronic prototyping, live coding, data sonification, looping and DJing.

As part of the 'Catalyst: Festival of Creativity' evaluation, a post-workshop questionnaire was circulated amongst attendees. Of 35 respondents, 68.6% reported some pre-existing familiarity with the workshop topic through their work or studies. However, participants' lack of confidence was clearly apparent in their self-rated knowledge of the topic on arrival to the workshop [34.3% not at all knowledgeable, 42.9% a little bit, 22.9% somewhat, 0% quite a bit, 0% a lot]. By the end of the session, these scores had improved in every case [overall: 0% not at all, 5.7% a little bit, 40% somewhat, 51.4% quite a bit, 2.9% a lot]. The majority reported living within Sheffield (63.6%) or its surrounds (6.1%), but almost a third (30.3%) had travelled from elsewhere in the UK to attend one of these workshops. Further, attendees fed back that they would recommend the workshop to others 'a lot' (94%), 'somewhat' (3%) and 'quite a bit' (3%). We therefore suggest that this community-building approach may provide a useful model for others to adopt in order to increase the participation of women and girls in sound and music technology in other localities.

*Sheffield-based network activities were supported by a 'Catalyst: Festival of Creativity' grant from Sheffield Hallam University.

Dr. A. V. Beeston is a KTP Associate in the Speech and Hearing Research Group in the Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello, Sheffield, S1 4DP (corresponding author e-mail: a.beeston@sheffield.ac.uk).

L. Cheesman is a Portfolio Officer in Business Support for the Faculty of Arts, Computing, Engineering & Sciences, Sheffield Hallam University, 2418 Harmer Building, Sheffield S1 1WB (e-mail: l.cheesman@shu.ac.uk).

Dr. E. D. Dobson is a Senior Lecturer in Music Technology, School of Music, Humanities & Media, University of Huddersfield, Huddersfield, HD1 3DH (e-mail: E.D.Dobson@hud.ac.uk).

REFERENCES

[1] G. Born and K. Devine (2015). Music technology, gender, and class: digitization, educational and social change in Britain. *Twentieth-Century Music*, 12(02), 135–172.

[2] Hu, X., Choi, K., Lee, J. H., Laplante, A., Hao, Y., Cunningham, S. J., & Downie, J. S. (2016). WiMIR - an informetric study on women authors in ISMIR. In *Proc. 17th International Society for Music Information Retrieval Conference*, 765–771.

[3] Freeman, J., Magerko, B., McKlin, T., Reilly, M., Permar, J., Summers, C. and Fruchter, E. (2014). Engaging underrepresented groups in high school introductory computing through computational remixing with EarSketch. In *Proc. 45th ACM Technical Symposium on Computer Science Education*, 85–90.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

The Relationship Between Emotion and Music Production Quality

David Ronan[1], Joshua D. Reiss[2] and Hatice Gunes[3]

[1*]Centre for Intelligent Sensing, Queen Mary University of London, UK, d.m.ronan@qmul.ac.uk
[2]Centre for Digital Music, Queen Mary University of London, UK
[3]Computer Laboratory, University of Cambridge, UK

***Abstract*— It is commonly known that music expresses emotion. In music production, the role of the mix engineer is to take a piece of recorded music and convey the emotions expressed as professionally sounding as possible. In this work, we investigated the relationship between music production quality and musically induced and perceived emotions.**

I. INTRODUCTION

There have been a number of studies that have looked at why people prefer some mixes over others. In one study, De Man et al. conducted a mixing experiment where groups of nine mix engineers were asked to mix 10 different songs [1, 2]. The mixes were then evaluated in a listening test to infer the quality, as perceived by a group of trained listeners. We are interested in the emotional impact of mix quality.

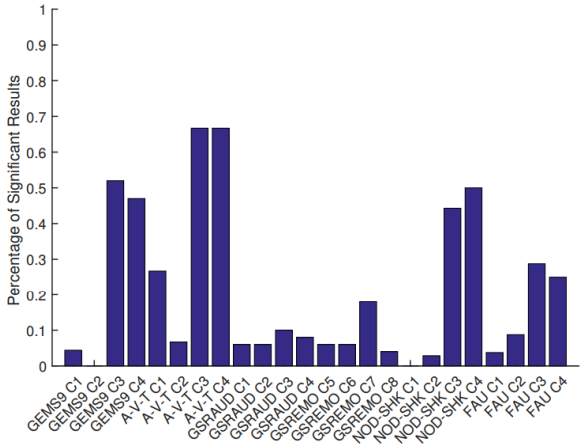
II. INTRODUCTION EXPERIMENTAL METHODS

We performed a listening test where 10 critical listeners and 10 non-critical listeners listened to 10 songs. There were two mixes of each song, the least preferred mix and the most preferred mix, where the mixes were rated for mix preference in a previous independent experiment [1, 2]. We measured each participant's subjective experience (GEMS-9 and Arousal-Valence-Tension), peripheral physiological changes (ECG and GSR), change in facial expressions and the number of head nods and shakes they made as they listened to each mix.

III. RESULTS AND DISCUSSION

By examining self-report scores, we showed that music production quality has more of an emotional impact on critical listeners. We also showed through facial expression analysis and head nod-shake detection that critical listeners have significantly different emotional responses to non-critical listeners for the most preferred mixes and to a lesser extent the least preferred mixes.

The results imply that the amount of emotion in a mix, whether it be perceived or felt, seems to matter the most to those with critical listening skills. This was most evident from the GEMS-9, Arousal-Valence-Tension, Head Nod/Shake Detection and Facial Action Unit results since they had the most amount of significant p-values. This can be clearly seen in Figure 1. If one was to take a cynical view on these results, it could be said that using the most professional and experienced mix engineer to mix a piece of music only really matters to those who have been trained to listen for mix defects, and mix quality has little bearing on the lay person emotionally. However, audio engineering could still be relevant to a listener without impacting them emotionally.



IV. CONCLUSION

We have shown that music production quality seems to only have a significant impact emotionally on those with critical listening skills.

ACKNOWLEDGMENTS

We would like to thank the (EPSRC) UK for funding this research.

REFERENCES

[1] B. De Man, M. Boerum, B. Leonard, R. King, G. Massenburg, and J. D. Reiss, "Perceptual evaluation of music mixing practices," in 138th Convention of the Audio Engineering Society, May 2015.

[2] B. De Man, B. Leonard, R. King, and J. D. Reiss, "An analysis and evaluation of audio features for multitrack music mixtures," in 15th International Society for Music Information Retrieval Conference (ISMIR 2014), October 2014.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Improved Pitch Trajectory Estimation for Polyphonic Single-Channel Audio Mixtures

Alejandro Delgado-Castro[1*] and John E. Szymanski[2]

[1*]Department of Electronics/Audio Lab, University of York, UK, adc533@york.ac.uk
[2]Department of Electronics/Audio Lab, University of York, UK.

***Abstract*— Estimating pitch trajectories for harmonic sources in single-channel polyphonic mixtures is a difficult task, especially when significant changes in volume and background instrumentation are present. A novel strategy is proposed to extract an improved and reliable dominant source pitch trajectory, based on one existing multipitch estimator, high-pass filtering, salience measurement and continuity.**

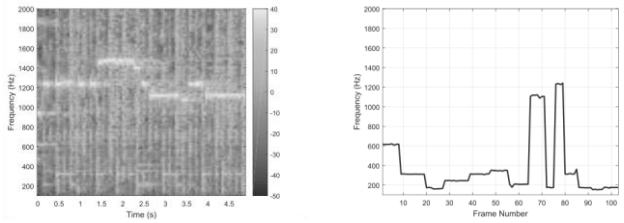
I. INTRODUCTION

Multipitch estimation, i.e. the task of extracting fundamental frequencies associated with harmonic sources in polyphonic signals, is an active area of research activity as an essential part of algorithms aimed towards applications such as automatic music transcription, audio source separation or creative sound transformation. It remains, however, a difficult endeavor, highly dependent on the number and relative volumes of the sources involved.

Duan's algorithm [1, 2] is an example of such a system. It estimates the number of harmonic sources present in the mixture and their fundamental frequencies, based on a maximum likelihood approach.

II. POTENTIAL PROBLEMS

Strong background instrumentation and changes in relative volume of the sources usually cause the multipitch estimator to deliver misleading pitch trajectories. In Figure 1, an incorrect estimation for a commercial recording of flute and drums is presented.



(a) (b)

Figure 1. (a) Spectrogram of the original audio mixture. (b) Mislead pitch trajectory estimated by Duan's algorithm [1, 2].

*Research supported by the University of Costa Rica (UCR) and the Costa Rican Ministry of Science, Technology and Telecommunications (MICITT).

A. Delgado Castro is a PhD student and J. E. Szymanski (john.szymanski@york.ac.uk) is senior lecturer within the Department of Electronics at the University of York. Their main research area is audio source separation applied to single-channel polyphonic signals.

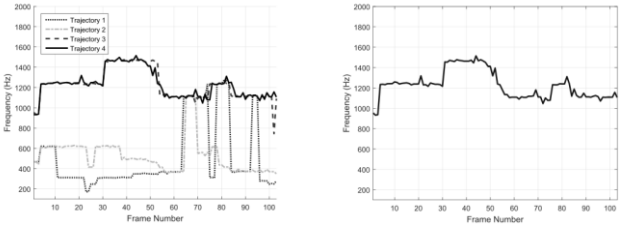
III. THE PROPOSED APPROACH

A novel strategy is proposed to extract an improved and reliable dominant pitch trajectory, as described below.

a) A set of high-pass filters is applied to the original mixture to create several modified versions of it.
b) Each filtered version is processed separately to generate a set of preliminary pitch trajectories.
c) A measure based on salience in frequency and continuity is used to select the best pitch candidate for each frame.

IV. RESULTS

Using four high-pass filters and the energy of the first five harmonic partial as a measure of salience, the results in Figure 2 were obtained.



(a) (b)

Figure 2. (a) Preliminary trajectories. (b) Final pitch trajectory.

V. CONCLUSIONS

The extraction of the dominant pitch trajectory can be improved by applying multipitch estimation to several versions of the original mixture and then evaluating the fundamental frequency estimates using different criteria.

ACKNOWLEDGMENT

The authors would like to thank the University of Costa Rica and the Costa Rican Ministry of Science, Technology and Telecommunications for their support in this research.

REFERENCES

[1] Z. Duan, B. Pardo, C. Zhang, "Multiple Fundamental Frequency Estimation by Modeling Spectral Peaks and Non-Peak Regions," *IEEE Transactions on Audio, Speech and Languages Processing*, vol. 18, no. 8, pp. 2121-2133, 2010.

[2] Z. Duan, "Multi-Pitch Analysis." `http://www.ece.rochester.edu/~zduan/multipitch/multipitch.html`.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Performable Spectral Synthesis via Low-Dimensional Modelling and Control Mapping

William Wilkinson[1*], Dan Stowell[1] and Joshua D. Reiss[1]

[1*]Centre for Digital Music, Queen Mary University of London, w.j.wilkinson@qmul.ac.uk

***Abstract*— Spectral modelling represents an audio signal as the sum of a finite number of partials – sinusoids tracked through sequential analysis frames. With the goal of real-time user-controllable synthesis in mind, we assume these observed partials to be correlated functions of time, and that there exists some lower-dimensional set of unobserved forcing functions driving the partials through a set of differential equations. Mapping of these unobserved functions to a user control space provides us with a hybrid approach to synthesis in which mechanistic controls are exposed to the user but the system's behavioural response to these mechanisms is learnt from data.**

I. INFERENCE

Spectral modelling synthesis [1] is capable of accurately reproducing a given audio recording by representing it as the sum of many sinusoids whose behaviour is tracked through time to create 'partials' [2]. In quasi-harmonic sounds, these partials can be thought of as functions of time and are likely to exhibit some shared behaviour.

We consider the maximum amplitude of the analysed partials and keep the most prominent, modelling the rest as a residual. We then assume that these partials are created by some low-dimensional set of forcing functions passing through a system of differential equations. Furthermore, if we choose this set of differential equations to be linear and define a joint Gaussian process prior [3] over the observed functions and the unobserved (latent) functions, we can infer from data both the differential equation parameters and the posterior distribution over the latent functions (Fig. 1). This approach provides a nonparametric way to estimate a low-dimensional representation of a system and its linear relationship to the full system, and is called latent force modelling [4,5].

II. SYNTHESIS

The inference process outlined above provides us with a system of differential equations with fixed parameters, and a low-dimensional set of forcing functions. Resynthesis could be performed by drawing samples from the posterior, but in order to enable performable control of the inferred system we have introduced an additional control mapping stage [6], in which the mean function of the Gaussian process posterior is replaced with some user input.

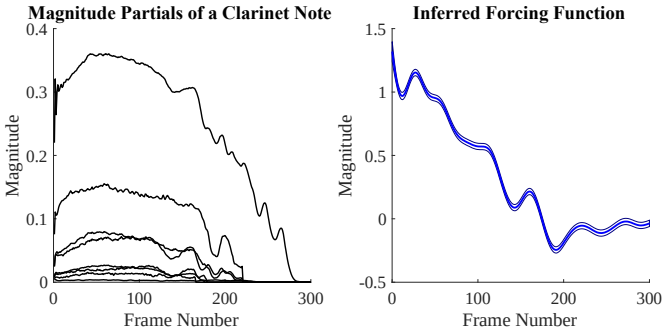


Figure 1. The magnitude over time of the first eight harmonics of a clarinet note (left) and the unobserved forcing function, plus its 95% confidence interval, learnt using the latent force model approach (right).

An experimental approach to defining this control mapping is taken. A performance gesture is recorded using MIDI CC messages which is then scaled and transformed (often by analysing its delta values) until it exhibits similar behaviour to the observed posterior mean. The scaling and transformation used then becomes the control mapping for all future gestures.

Once the control mapping has been defined, the system is implemented in real-time by interpolating frame-level data to obtain sample-level data and applying the learnt mappings to a series of sinusoidal oscillators representing the partials. Implementations are built as VST instruments with MIDI input using C++ and the JUCE framework.

REFERENCES

[1] X. Serra and J. O. Smith. "Spectral modelling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition." *Computer Music Journal*, pp. 12-24, 1990.

[2] M. Klingbeil. "Software for spectral analysis, editing, and synthesis." *Proceedings of the International Computer Music Conference.* 2005.

[3] C. E. Rasmussen. "Gaussian processes for machine learning." 2006.

[4] M. A. Alvarez, D. Luengo and N. D. Lawrence. "Latent force models." *AISTATS. Vol. 12*, 2009.

[5] J. Hartikainen and S. Sarkka. "Sequential inference for latent force models." *Proc. 27th Conf. Uncertainty in Artificial Intelligence*, pp. 311-318, 2011.

[6] P. Popp and M. Wright. "Intuitive real-time control of spectral model synthesis." *NIME.* 2011.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Understanding Creativity and Autonomy in Music Performance and Composition: A proposed 'toolkit' for research and design

Alan Chamberlain[1*], David De Roure[2], Pip Willcox[3], Steve Benford[1], Chris Greenhalgh[1]

[1*]Department of Computer Science, MRL, University of Nottingham, UK, azc@cs.nott.ac.uk
[2]OeRC [3]Centre for Digital Scholarship, University of Oxford, UK

***Abstract* - This paper proposes the bringing together of a series of tools and methods in order to understand the interplay and interaction between human creativity and 'autonomous' algorithmic systems in respect to performing and composing music. The aim of the paper is to prompt a discussion within the digital music research community, and associated disciplines in order to develop an understanding of some of the issues and debates relating to the area of research and the design of autonomous systems.**

I. TOOLS & METHODS

Here we outline a set of developing tools and research methodologies that could help us in further understanding the use and application of autonomous systems in a real world setting. We envisage a set of tools that together could support the composition, and performance of music, and also allow researchers/performers to document their activities, methods and ideas. We propose an initial system would allow people to interact and create content in a number of different ways, using a range of techniques. Capturing this and documenting it are also key to the proposal. We use the term *a*Sketch (autonomy sketch) to define the materials that are finally brought together to support and define the way that autonomous systems are used, the way that they support creative practices and open up further design possibilities. The *a*Sketch will enable the creation of a Digital Music Object that will encapsulate data from the systems and methods we list below. *Algorithmic Content Generation* – Numbers into Notes [1] we would aim to use this as a tool to demonstrate, give people an understanding of algorithmic content generation and create content (midi output etc.) A series of examples, performances and a web-based tool exist. An Arduino-based system is in progress. *Code-based Triggering Systems* - Muzicodes [2] (audio), Artcodes [3] (visual). Both of these systems are able to trigger content, in particular audio triggering can be *specific*, triggered by a certain set of notes/rhythm, or can be made more *general* and can be triggered randomly – perhaps by audio in the environment. A*daptable Visual Interfaces* – As part of the system we seek to develop methods of visually interfacing with the Algorithmic Content Generation tools so that we could see how performers, composers and designers envisage tools for autonomous music generation, beyond numerical representation. *Sensor-based systems* – Sensors would be used to support the creation of content. These could be as simple as microphones and may be used for input, adaption and recording (*data*). It may be that the data recorded from a given part of a performance is later fed back into the performance, or passed on to other performers/composers. *Trajectory-based* [3] *Mapping and Documentation* – These tools will allow users and researchers to map out their trajectories as they use different tools and make different decisions. In the future we would hope to integrate *Mechanisms for Collaboration*, and *Adaptable Physical Interfacing* that could link to mechanical systems.

Methodologically we would take a none-theoretical auto-ethnography [4][5] based approach in order that we could fully take advantage of the user's understanding and approach to the system. The research would be practice-led (artifact creation) and practice-based (knowledge creation) – and participatory in that the methodology would create a mechanism by which people could have input into the research/system, as a member of the research team and as someone using the system. We would also encourage researchers from different disciplines that may have different research agenda. This may further add to our understanding of the interplay between performer/composer and the role of autonomy in the system. To conclude, understanding the interaction between people and autonomous systems in respect to music and creativity is complex, theoretical and can be difficult to articulate. We would hope that that this initial foray into the autonomous systems and human creativity in relation to musical performance/composition might serve as a provocation, and as such a platform to engender debate and discussion across disciplines.

ACKNOWLEDGMENT

This research was supported through the following EPSRC project: Fusing Semantic and Audio Technologies for Intelligent Music Production and Consumption (EP/L019981/1) & Transforming Musicology, funded by the UK Arts and Humanities Research Council (AHRC) under grant AH/L006820/1 in the Digital Transformations programme.

REFERENCES

[1] De Roure, David. Numbers Into Notes: http://demeter.oerc.ox.ac.uk/NumbersIntoNotes/.

[2] Greenhalgh, C., Benford, S, & Hazzard, A (2016) ^muzicode$: composing and performing musical codes. In: Audio Mostly 2016, 4-6 Oct 2016, Norrköping, Sweden.

[3] Steve Benford, Adrian Hazzard, Alan Chamberlain, Kevin Glover, Chris Greenhalgh, Liming Xu, Michaela Hoare, Dimitrios D. (2016) "Accountable Artefacts: the Case of the Carolan Guitar", Proceedings of CHI'16, May 07 - 12, San Jose, CA, USA, ACM

[4] Benford, S., Giannachi, G., Koleva, B. and Rodden, T., (2009). From interaction to trajectories: designing coherent journeys through user experiences In: Proceedings of CHI '09, ACM.

[5] Parisa Eslambolchilar, Mads Bødker and Alan Chamberlain (2016) "Ways of Walking: Understanding Walking's Implications for the Design of Handheld Technology via a Humanistic Ethnographic Approach". *in* Human Technology: An Interdisciplinary Journal on Humans in ICT Environments.

[6] Mads Bødker and Alan Chamberlain (2016) "Affect Theory and Autoethnography in Ordinary Information Systems", Proceedings of 2016 European Conference on Information Systems, ECIS 2016.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Automatic Control of the Dynamic Range Compressor Using a Regression Model and a Reference Sound

Di Sheng[1*] and György Fazekas[2]

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

***Abstract*— Controlling audio effects requires substantial professional knowledge of music production and mixing. However, using audio effects is important for processing all kinds of audio materials and it should not be an exclusive skill for professional audio engineers. This paper provides a novel control method for audio effects. Instead of using parameters (e.g. threshold at 30db) or semantic terms (e.g. a bright EQ), it uses another audio file as "description" and applies an audio effect to the user's file to get the desired sound that is closer to the example in terms of perceptually relevant production parameters.**

I. INTRODUCTION

Production is an important part of the music industry. Research on intelligent control of audio effects can significantly reduce the labour and time that the production process takes. In this paper, we propose a tool that is able to configure the audio effect parameters using a reference track.

The proposed intelligent tool requires a considerable effort to realise. As a starting point for our work, the task is simplified using several assumptions. We only consider simple audio materials, notes and mono-timbral loops. The only audio effect considered is the dynamic compressor. The proposed solution has the following components: 1) an audio feature generator that generates features corresponding to each parameter, 2) a regression model that maps from audio features to audio effect parameters, 3) a predictor that corresponds to this model, and finally 4) a simple method for computing perceptual audio similarity to compare the result from the predictor and the reference note.

II. METHOD

We use linear regression to model the relationship between low-level audio features and compressor parameters. To be more specific, four of the most important compressor parameters are considered: threshold, ratio, attack and release time. A selected group of six high order statistic features are used for threshold/ratio, and in addition to this, four temporal features are used to estimate attack and release time. In this research, we use a database containing 60 different violin notes. Four audio datasets are generated out of these notes corresponding to individual parameters, and these are used for further prediction.

*Research supported by Queen Mary University London.

Di Sheng, György Fazekas are with Queen Mary University London, Center for Digital Music.

The workflow of this research can be separated into the following processing and evaluation steps:

1. Extracting features and forming regression models;
2. Examine model efficiency by using cross-validation;
3. Using the predicted parameters from a reference note to process a raw note (from another source) and testing audio similarity of the output with the reference (both when the origin of the reference note is available and when it is not).
4. Using the predicted parameter from a reference mono-timbral loop to process a raw loop and testing similarity of the output with the reference (in both cases as 3).

The audio similarity measurement in step 3 and 4 is based on computing the variational Bayes approximation of the KL Divergence [2] between Gaussian Mixture Models fitted on MFCC features [1] extracted from the audio.

III. RESULTS AND CONCLUSION

The results of the series of experiments show a promising trend: after applying the predicted parameters, the similarity with the reference is higher than its origin (the unprocessed audio). The following table shows the result from the 4th step, where N_{pro} is the reference note, T is the output of the algorithm, and R is its origin. Although, the results are promising, further study is required and the algorithm needs to be improved in multiple aspects.

TABLE I. SIMILARITY BETWEEN NOTES

	$D(N_{pro},R)$	$D(N_{pro},T)$
Threshold	52.05	24.96
Ratio	63.86	25.43
Attack	51.61	24.60
Release	61.33	24.48

REFERENCES

[1] Beth Logan and Ariel Salomon, "A music similarity function based on signal analysis.," in ICME, 2001.
[2] J R Hershey and P A Olsen, "Approximating the kullback leibler divergence between gaussian mixture models," in 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07. IEEE, 2007, vol. 4, pp. IV–317.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Computational Corpus Analysis of Native American Music

Kerstin Neubarth[1], Daniel Shanahan[2] and Darrell Conklin[3,4]

[1]Canterbury Christ Church University, Canterbury, United Kingdom
[2]Louisiana State University, Baton Rouge, United States
[3]University of the Basque Country UPV/EHU, San Sebastian, Spain
[4]IKERBASQUE: Basque Foundation for Science, Bilbao, Spain

***Abstract*— Building on Frances Densmore's collection and analysis of North American Native music, the current study applies supervised descriptive data mining to a corpus of 2218 Native American songs in order to identify distinctive global-feature patterns which characterise songs of different indigenous tribes. Discovered patterns are related to Densmore's original findings.**

I. INTRODUCTION

In 1954 ethnomusicologist Bruno Nettl observed about indigenous music of the North American continent that "more musical material is available from this large area […] than from any other of similar size" [3, p. 45]. Samples of Native American music have been the subject of corpus-level analysis including comparative studies [2], large-scale surveys [3] and statistical tests [1], or have formed part of cross-cultural research [5]. Many studies have used material collected by anthropologist Frances Densmore (1867-1957). For a large part of her collection, Densmore herself presented quantitative analyses identifying common and distinctive musical features in the repertoires of different tribes, manually extracting features from the recorded and transcribed songs, determining occurrence frequencies of features, and comparing the proportional frequencies between tribes [7]. In effect, these analyses provide an early, manually executed, example of supervised descriptive pattern discovery [4]. In the current work we apply computational methods to discover distinctive patterns.

TABLE I. SELECTED DENSMORE ATTRIBUTES

Attribute	*Description*
Tonality	tonality: third above keynote
FirstReKey	first note relative to keynote
LastReKey	last note relative to keynote
LastReCompass	last note relative to compass of song
Compass	number of tones comprising compass
Material	tone material
Accidentals	accidentals: chromatic alterations of tones
Structure	relation between contiguous accented tones
FirstDir	direction of first progression
FirstMetre	part of measure on which song begins
InitMetre	rhythm (metre) of first measure
MetreChange	change of time (measure-lengths)
RhythmicUnits	rhythmic unit(s) of song

II. DATA AND ANALYSIS

The analysis corpus of this study currently consists of 2218 songs organised into 16 tribes [6]. For 1760 songs Densmore provides tabular analyses by global attributes (Table 1). From these analyses we collated Densmore's features, cleaned duplicate features and harmonised inconsistent values.

Supervised descriptive pattern mining [4] was applied to discover distinctive global-feature patterns, i.e. conjunctions of song-level attribute-value pairs which are significantly over-represented in a target group relative to the background groups.

III. RESULTS

Computationally discovered patterns can be related to Densmore's analyses in four ways: (1) statistically significant features corresponding to observations by Densmore; (2) features highlighted by Densmore which in our analysis are distinctive but the distinctiveness is not statistically significant; (3) significantly distinctive features different from features suggested by Densmore; and (4) statistically significant combinations of features individually analysed by Densmore.

REFERENCES

[1] R. H. Gundlach. A quantitative analysis of Indian music. *The American Journal of Psychology*, 44(1), 1932, pp. 133–145.
[2] G. Herzog. A comparison of Pueblo and Pima musical styles. *The Journal of American Folklore*, 49(194), 1936, pp. 283–417.
[3] B. Nettl. North American Indian musical styles. *The Journal of American Folklore*, 67(263, 265, 266), 1954, pp. 44–56, 297–307, 351–368.
[4] P. K. Novak, N. Lavrač, and G. Webb. Supervised descriptive rule discovery: a unifying survey of contrast set, emerging pattern and subgroup mining. *Journal of Machine Learning Research*, 10, 2009, pp. 377–403.
[5] P. E. Savage, S. Brown, E. Sakai, and T. E. Currie. Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences*, 112(29), 2015, pp. 8987–8992.
[6] D. Shanahan, K. Neubarth, and D. Conklin. Mining musical traits of social functions in Native American music. *Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR 2016)*, New York, USA, 2016, pp. 681–687.
[7] Smithsonian Institution. *List of Publications of the Bureau of American Ethnology*. Washington, DC: Smithsonian Institution BAE Bulletin 200, 1971.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Towards a Music Language Model for Audio Analysis

Adrien Ycart[1*] and Emmanouil Benetos[1]

[1*] C4DM, Queen Mary University of London, {a.ycart, emmanouil.benetos}@qmul.ac.uk

***Abstract*— Polyphonic Automatic Music Transcription remains a challenging problem. Many studies focus on the extraction of features from audio signals; we focus here on Music Language Models that help turn those features into a musically meaningful piano-roll representation. We review related work on music language models, and propose ways to improve it.**

I. INTRODUCTION

Automatic music transcription (AMT) is one of the most widely discussed topics in Music Information Retrieval. Still, it remains a challenging task, in particular in the case of polyphonic music.

In a similar fashion to natural speech transcription, where an *acoustic model* extracts features from an audio signal and a *language model* uses high-level knowledge to produce the actual transcription, AMT is performed by estimating at each time frame the pitches that compose an audio recording, usually via some kind of time-frequency representation, and then linking the time-frame estimations using some musical knowledge to build a binary piano-roll representation. While the former task has been widely discussed in the literature, the latter has received little attention until quite recently.

We propose to develop a *Music Language Model* (MLM) that would encode the prior musical knowledge we have on the output we want to obtain, in order to improve AMT.

II. RELATED WORK

MLMs have not been used specifically in audio analysis since quite recently. Temperley [1] was one of the first to propose a joint model for harmony, rhythm and stream separation, and suggested its use for transcription. The model is quite comprehensive; however, it seems to be intractable in real-life applications.

More recently, Raczyński et al. [2] have designed a probabilistic model of harmony to post-process the output of a multi-pitch NMF estimator, but without considering temporal aspects (long-term structure, rhythm).

Other approaches have used neural networks to post-process the output of an acoustic model. Boulanger-Lewandowski proposed in [3] a neural network that estimates a pitch distribution at each time-step given the previous ones, and the observation given by an acoustic model. However the two models are chained, thus errors are propagated. Some further developments have attempted to integrate the two models in a single neural network [4], but here again, rhythm and long-term structure are neglected.

Many other studies have focused on MLMs. Lerdahl & Jackendoff [5] proposed a founding model for tonal music, which was implemented in [6]. The IDyOM model [7] also models melodic expectation from a cognitive point of view. However, none of these models were intended to be integrated in an audio analysis system.

III. HYPOTHESES AND DEVELOPMENTS

We focus our study on Western tonal music. We assume that understanding the temporal structure, both short-term (rhythm, meter) and long-term, is necessary to correctly model it. The lack of data to infer model parameters is also a big challenge, which leads to over-simplified models. We assume that inputting explicit knowledge might reduce the quantity of data needed for training. We also assume that the users of transcription systems are experts and are able to correct possible errors in the system output, which allows us to learn from their corrections on top of a corpus. We suggest that some errors are more significant than others, thus our model should take extra care to remove them, in order to preserve the musical content conveyed in the original recording. We also believe that the acoustic model and the MLM benefit from each other's knowledge, thus that a successful AMT system should join the two models instead of chaining them.

*A. Ycart is supported by a QMUL EECS Research Studentship. E. Benetos is supported by a RAEng Research Fellowship (RF/128).

REFERENCES

[1] Temperley, D. (2009). A Unified Probabilistic Model for Polyphonic Music Analysis. *Journal of New Music Research*, *38*(1), 3–18.

[2] Raczyński, S. A., Vincent, E., & Sagayama, S. (2013). Dynamic Bayesian networks for symbolic polyhonic pitch modeling. *IEEE Transactions on Audio, Speech, and Language Processing*, *21*(9), 1830–1840.

[3] Boulanger-Lewandowski, N., Vincent, P., & Bengio, Y. (2012). Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription. *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, (Cd), 1159–1166.

[4] Sigtia, S., Benetos, E., & Dixon, S. (2015). An end-to-end neural network for polyphonic music transcription. *IEEE Transactions on Audio, Speech, and Language Processing*, *24*(5), 927–939.

[5] Lerdahl, F., & Jackendoff, R. (1983). A Generative Theory of Tonal Music. *Musicology Australia*, *9*(1), 72–73.

[6] Hirata, K., Tojo, S., & Hamanaka, M. (2006). Implementing "A Generative Theory of Tonal Music." *Journal of New Music Research*, *35*(4), 249–277.

[7] Pearce, M. T. (2005). *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Designing a Highly Expressive Algorithmic Music Composition System for Non-Programmers

Matt Bellingham[1*], Simon Holland[2] and Paul Mulholland[3]

[1*]Department of Music, University of Wolverhampton, UK, matt.bellingham@wlv.ac.uk
[2]Music Computing Lab, Centre for Research in Computing, The Open University, UK
[3]Knowledge Media Institute, Centre for Research in Computing, The Open University, UK

***Abstract*—Algorithmic composition systems allow for the partial or total automation of music composition by formal, computational means. Typical algorithmic composition systems generate nondeterministic music, meaning that multiple musical outcomes can result from the same algorithm - consequently the output is generally different each time the algorithm runs.**

Here we present an algorithmic composition system designed to meet the needs of a particular user group: undergraduate Music Technology students. Unexpectedly, the specific needs of this user group led us to radical design decisions, which ended up reshaping the fundamentals of the underlying programming language design. Our users are typically not programmers, and they are often not traditional musicians. While they may be conversant with some elements of music theory, their background is often as self-taught music producers with experience of making music electronically using music sequencers/DAWs.

There are many existing tools which can be used for algorithmic music composition, but from the perspective of our target user group, they all exhibit various limitations [1, 2]. For example, Ableton Live offers a simple and efficient UI which provides a number of tools to assist loop-based composition, but the software lacks the expressivity and generality required for algorithmic work. Graphical programming languages such as Max and Pure Data are highly expressive but require the user to have sufficient pre-existing musical knowledge to build their own musical structure into patches. Text-oriented languages such as SuperCollider and Sonic Pi offer high expressivity and are well adapted to musical structures, but require the user to manipulate musical materials through structures and syntax shaped primarily by the concerns of conventional programming languages.

Our system is designed to meet several needs: to enable our target users to create algorithmic music of arbitrary complexity; to facilitate graphical programming with minimal syntactical concerns; and to make common musical tasks simple. As a by-product, these properties promote learning the concepts of algorithmic composition via hands-on experience.

In the new algorithmic composition system, the principal programming primitive is called a *chooser*. Every instance of this primitive affords such musically useful actions as: loop X until Y finishes; hierarchical organization; random selection; and choice points. Due to this and related design decisions, the system has relatively low viscosity and low verbosity: i.e. musically desirable changes are relatively easy to make, and musically complex constructs can be expressed concisely. These properties derive from what is technically known as the 'closeness of mapping' of the notation, i.e. how closely the notation corresponds to the problem world [3]. In part, this is achieved because the programming language gives the affordances of the graphical track/mixer view of a sequencer/DAW while providing the full power of a recursive programming language.

At present the design is being iteratively refined using the programming walkthrough method [4] and implemented using SuperCollider as the back end.

REFERENCES

[1] Bellingham, M., Holland, S. and Mulholland, P. (2014a) A Cognitive Dimensions analysis of interaction design for algorithmic composition software, In *Proceedings of Psychology of Programming Interest Group Annual Conference 2014*, du Boulay, B. and Good, J. (eds.), University of Sussex, pp. 135–140, [online] Available from: http://www.sussex.ac.uk/Users/bend/ppig2014/15ppig2014_submission_10.pdf.

[2] Bellingham, M., Holland, S. and Mulholland, P. (2014b) *An analysis of algorithmic composition interaction design with reference to cognitive dimensions*, The Open University, [online] Available from: http://computing-reports.open.ac.uk/2014/TR2014-04.pdf.

[3] Green, T. R. and Petre, M. (1996) Usability Analysis of Visual Programming Environments: a `cognitive dimensions' framework, *Journal of Visual Languages and Computing*, 7, pp. 131–174.

[4] Bell, B., Citrin, W. V., Lewis, C. and Rieman, J. (1992) The Programming Walkthrough: A Structured Method for Assessing the Writability of Programming Languages; CU-CS-577-92, *Computer Science Technical Reports*, Paper 554, [online] Available from: http://scholar.colorado.edu/csci_techreports/554.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Explaining Predictions of Machine Listening Systems

Saumitra Mishra[1], Bob L. Sturm[1] and Simon Dixon[1]

[1]Centre for Digital Music, Queen Mary University of London, UK, saumitra.mishra@qmul.ac.uk

***Abstract*— We adapt local, interpretable and model-agnostic explanations [1] for use with a machine listening system, and demonstrate it for singing voice detection. Such explanations provide ways to understand the behaviour of machine listening systems and to judge their generalisation. One approach works in the time-domain, and the other in the frequency-domain.**

I. MOTIVATION

Explaining a prediction of a machine listening system in an interpretable way can help to understand it. Interpretability is essential as often the explanations learned on the feature representations are too complex to highlight the mapping between the instance space and the classifier predictions. Consider the simple binary decision tree (BDT) system in Fig. 1. The input is a *T*-msec segment of audio, from which the system extracts *D* Mel-frequency cepstral coefficients (MFCC). The system labels the feature as either class "A" or class "B" based on thresholds learned via training.

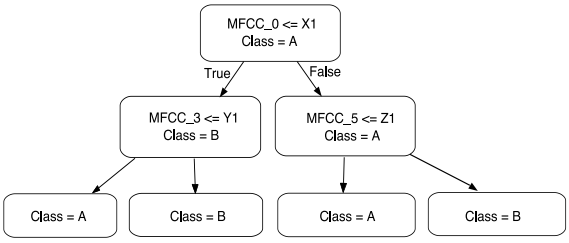


Figure 1. A simple binary decision tree used for classifying audio.

A BDT seems transparent because one can trace the cause of a particular classification, in this case in terms of the MFCC coefficients and thresholds. This transparency, however, is hard to interpret due to the obscure nature between the MFCC coefficients and perception (aside from the *zeroth MFCC*, which is related to the energy of an input). This makes it difficult to relate system behaviour to input audible qualities, and so motivates using "interpretable representations" for explaining system behaviour.

II. INTERPRETABLE EXPLANATIONS FOR MACHINE LISTENING

We adapt and apply the "local interpretable model-agnostic explanation" (LIME) algorithm [1] to machine listening. LIME explains a prediction by approximating the decision boundary near an instance by an interpretable model (e.g. sparse linear model, decision tree) learnt on interpretable data representations (e.g. words in an e-mail, super-pixels in an image). Our extension, Sound-LIME (SLIME) [2], proposes two kinds of interpretable representations: one based on temporal segmentation (super-samples, or audio segments), and the other based on spectral bins.

We apply SLIME to a system for singing voice detection [3]. The system extracts 13 MFCC coefficients from a 1-sec duration instance (frame), and uses a binary decision tree built from 22720 frames. The system achieves an overall accuracy of 82.38% on a hold-out dataset. SLIME generates explanations for each prediction of the BDT by uniformly segmenting each instance into ten "super-samples" (indexed 0-9), building a linear model, and selecting the top three super-samples that best explain the instance prediction. Table 1 lists these explanations for five instances.

TABLE I. SLIME EXPLANATIONS FOR 5 INSTANCES. COLUMNS ARE: TRUE SINGING DURATION; BDT PROBABILITY THAT INSTANCE IS VOCAL; SUPER-SAMPLES EXPLAINING WHY BDT 'BELIEVES' INSTANCE IS VOICE (SUPER-SAMPLES WITH VOICE IN PARENS).

Instance ID	*True Singing Dur. (msec)*	*P-Vocal*	*Predicted (True) Super-Sample*
5	0	0.95	5,9,6 (None)
38	100	0.95	6,5,7 (9)
39	600	0.64	1,0,5 (4-9)
62	800	0.64	3,9,2 (0-7)
63	300	0.64	2,3,6 (0-2)

Table 1 lists super-samples identified by SLIME in order of decreasing influence on predictions of the BDT. For example, the BDT is 64% confident that instance 63 has 'singing voice'. SLIME finds super-samples 2, 3 & 6 have a strong influence on this confidence. When we listen to the super-samples we find irregularities in the system behaviour. For example, the model labels instance 38 'signing voice' with 95% confidence, but listening to the super-samples reveals them to not be of voice. The explanations generated by SLIME thus suggest that in spite of having high classification accuracy, the system may not be trustworthy for detecting singing voice.

REFERENCES

[1] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?": Explaining the predictions of any classifier," in *Proc. KDD*, 2016.

[2] S. Mishra, B. L. Sturm, and S. Dixon, "Local interpretable model-agnostic explanations for machine listening systems," in *Proc. ICASSP* (submitted), 2017.

[3] B. Lehner, R. Sonnleitner, and G. Widmer, "Towards light-weight, real-time-capable singing voice detection," in *Proc. ISMIR*, 2013.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Experimental Digital Humanities: Creative interventions in algorithmic composition on a hypothetical mechanical computer

David De Roure[1*] Pip Willcox[2] Alan Chamberlain[3]

[1*]OeRC [2]Centre for Digital Scholarship, University of Oxford, UK, david.deroure@oerc.ox.ac.uk
[3]Department of Computer Science, Mixed Reality Lab, University of Nottingham, UK

Abstract **- Ada Lovelace wrote of Charles Babbage's hypothetical room-sized steam-powered Analytical Engine: "*Supposing, for instance, that the fundamental relations of pitched sounds in the science of harmony and of musical composition were susceptible of such expression and adaptations, the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent*." [1]. Supposing Charles Babbage had built the Analytical Engine and Ada Lovelace had pursued its potential for musical composition, what music might have she have created?**

I. MUSIC, MATHS & METHODOLOGY

The Numbers into Notes project, in its first phase, explored how this might have occurred nearly two centuries ago using the mathematics of the time. Our online tool can be used to generate musical fragments, which are the basis of compositions today, and was used at an 'Ada sketches' performance [2] by composer *Emily Howard* and mathematician *Lasse Rempe-Gillen* at the Royal Northern College of Music, in which the audience generated music for human performers. Now we are asking what Lovelace might do today, with a microcontroller instead of the Analytical Engine. In our explorations, we have also programmed 'devices' to generate music, with creative interventions by humans to compose and influence the experience in locative sound. Thus we explore embodied interaction beyond the bounds of the original approach. Our work explores the relationship between autonomy, algorithm, and human creativity in music creation, using a methodology of digitally enabled "experimental humanities" akin to the material methodologies used in experimental archaeology. However, our 'radical' approach brings proposed technologies into a contemporary setting, siting them within current academic discourses, in this case relating to human creativity, algorithms and autonomous systems. We argue that looking back in time at the context of a given work is not the only way to understand it, but an abreaction of the work, in our case, under the *influence of a given set of contemporary technologies*, is also of value, in particular when one is able to use evaluative methodologies adapted from HCI (Human-Computer Interaction) design, which take a more auto-ethnographic [3] stance. This enables us to gain a fuller appreciation of the interplay and situated nature of creativity and autonomy as based in tools, provoked by and reimagined from a historical context, which brings mathematics, music, technology and the humanities together in a form that arguably engenders creativity and disruptive understandings of the past through a contemporary technological lens. An emergent debate is focused around humans, algorithmic content generation and the (*perceived*) role of autonomy - *in music-based practice*. Based on our multi-disciplinary research, we have started to develop tools that could be developed to support and further understand music-based practices; such as creativity, composition and improvisational performance, and the role autonomous systems as understood though the lens of human computer interaction (HCI) in such contexts. Our aim is to develop a toolkit that could be used to enable deeper understanding of "*the relationship between autonomy, algorithm, and human creativity in music creation*" and that could further develop techno-humanities-based research. Bringing together systems such as *Numbers into Notes* [4], *Muzicodes*, sensor-based systems, and a trajectory-based 'notation' system, would afford us and other users the opportunity to sketch ideas, and document and map the interactions and interplay between autonomy and creativity. To conclude, historical re-imaginings can bring prospective, imagined, theoretical technologies into a contemporary setting, and in so doing, engender new responses and discourses, which, in our case relate to music practice (performance, composition and production) and the philosophy of technology. They have enabled us to think about the nature of a more 'experimental humanities'. These interpretations of the past, using contemporary technologies and concepts framed in producing and performing music, lead us to question, understand and push the boundaries of what might be possible in understanding creativity and autonomy.

ACKNOWLEDGEMENT

This research was supported through the following EPSRC project: Fusing Semantic and Audio Technologies for Intelligent Music Production and Consumption (EP/L019981/1) & Transforming Musicology, funded by the UK Arts and Humanities Research Council (AHRC) under grant AH/L006820/1 in the Digital Transformations programme.

REFERENCES

[1] Lovelace, A.A. (Trans.) (1843) Sketch of the analytical engine invented by Charles Babbage, Esq. with notes by the translator. Scientific Memoirs, 3, 666-731.
[2] Science Lates—Ada sketches, 27 July 2016: http://manchestersciencecity.com/visit/event/science-lates-ada-sketches/.
[3] Mads Bødker and Alan Chamberlain (2016) "Affect Theory and Autoethnography in Ordinary Information Systems", Proceedings of 2016 European Conference on Information Systems, ECIS 2016, AIS
[4] De Roure, David. Numbers Into Notes: http://demeter.oerc.ox.ac.uk/NumbersIntoNotes/.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Perceptual Evaluation of Synthesised Sound Effects

David Moffat[1*] and Joshua D. Reiss[1]

[1*]Center for Digital Music, Queen Mary University of London, UK, d.j.moffat@qmul.ac.uk

***Abstract*— Sound synthesis is the method of artificially producing sounds. A range of sound effects were generated through the use of five different synthesis techniques, across eight sound classes to produce sixty six different example audio samples. Perceptual evaluation was performed to identify the perceived realism of each of the synthesis techniques. The results show that some synthesis techniques can be considered as realistic as a recorded sample, when listening directly to audio samples.**

I. INTRODUCTION

Sound Synthesis (SS) is the technique of generating sound through artificial means. Synthesis has been demonstrated in the area of sound effect, which can be used in production of a range of popular media, such as video games, TV, film and augmented or virtual reality. A listening test was undertaken in which participants were asked to rank audio files in terms of their perceived realism.

II. METHOD

Synthesis of a range of sound effects was performed through a range of methods including statistical modelling, sinusoidal modelling, sinusoidal modelling with residual, physically informed modelling, granular synthesis, concatenative synthesis and additive synthesis.

Thirteen participants were asked to evaluate sounds for eight categories (applause; babble; bees; fire; rain; stream; waves; wind) and in every category between six and eleven samples were provided. Sixty six samples were evaluated in total. All samples were loudness normalised. Each category had at least some anchor and at least one recorded sample. The recorded samples were all select by a group of five experienced critical listeners as being realistic samples. Participants were asked to rate all samples within a given category, relative to all the other samples within the category. They were to rate the samples relative to how realistic they perceived the sounds on a continuous linear scale labelled from "very unrealistic" through "quite unrealistic", "quite realistic" to "very realistic". The provided synthesised samples were also compared to at least one reference and at least one anchor per category, similar to a mushra style test. Participants did not have any information regarding the samples, other than a reference number, which was randomised. The initial locations of samples were randomised and the order of categories participants were evaluating was randomised. The listening test was constructed using the Web Audio Evaluation Toolbox [1].

III. RESULTS

ANOVA was undertaken to determine the if the distributions of the results are the same. SPAD [2] was the only synthesis method that cannot be considered statistically significant from the reference audio samples. It can be considered as realistic as a recorded sample in the case of all sound classes. The user ratings are presented in Figure 1. The red line represents the median. The end of the notches represents the 95% confidence intervals, and the end of the boxes represent the 1st and 3rd quartiles. The end of the whiskers represent the range of the data not considered as an outlier. Red crosses are outliers.

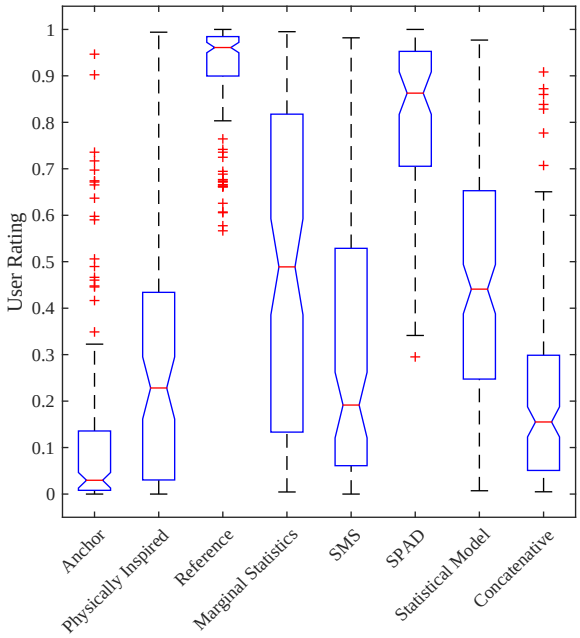


Figure 1. Perceieved Realism of all Synthesis Methods

REFERENCES

[1] N Jillings, B De Man, D Moffat and JD Reiss, "Web Audio Evaluation Tool: A browser-based listening test environment" in Proc. International Sound and Music Computing Conference (2015).

[2] Verron, C., Aramaki, M., Kronland-Martinet, R., & Pallone, G. (2010). A 3-D immersive synthesizer for environmental sounds. IEEE Transactions on Audio, Speech, and Language Processing, 18(6), 1550-1561.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Comparative Evaluation of Rhythm Transcription Algorithms on Polyphonic Piano Datasets

Eita Nakamura[1,2] and Kazuyoshi Yoshii[1]

[1]Graduate School of Informatics, Kyoto University, Japan
[2]Centre for Digital Music, Queen Mary University of London, UK

***Abstract*— This study presents comparative evaluations of rhythm transcription algorithms, which convert polyphonic MIDI performances into music scores. Seven algorithms including the most recent ones are compared on two datasets of expressive polyphonic piano performances; one containing performances with polyrhythms and the other containing standard polyphonic performances. We found that methods based on statistical learning tends to have better accuracies and a method using an HMM-based model of multiple-voice structure had the best accuracy for polyrhythmic performances.**

Music transcription is a fundamental problem in music information processing, requiring the extraction of pitch and rhythm information from music audio signals. There have been many studies on converting a music audio signal into a piano-roll representation based on acoustic modelling of musical sound (see [1,2] for reviews). To obtain a music score, we must recognise quantised note lengths (or note values) of the musical notes in piano rolls, which is called rhythm transcription [3–9].

In early studies on rhythm transcription, methods based on the connectionist approach [3] and the rule-based approach [6] were studied. Since around 2000, methods based on statistical models have been widely studied. Takeda et al. proposed a method using a Markov model of note values [5] and Raphael used a Markov process defined on a grid of beat positions [6]. Extending his previous study [4], Temperley developed a model unifying metrical analysis, harmony analysis and stream segregation [7]. To capture the multi-stream structure of polyphonic music, models based on probabilistic context-free grammar (PCFG) [8] and an extension of hidden Markov model (HMM) [9] have been proposed recently.

Despite the importance of the subject, there have been no reports on the comparison of different rhythm transcription algorithms. In this study we directly compare seven representative algorithms reviewed in the previous paragraph and discuss the results. Two datasets, "polyrhythmic dataset" and "standard polyphony dataset," were prepared for this.

*Research supported by KAKENHI Nos. 24220006, 26280089, 26700020, 15K16054, 16H01744 and 16J05486, JST OngaACCEL Project and JSPS Research Fellowship.

E. Nakamura is with Kyoto University, Kyoto 606-8501, Japan (corresponding author e-mail: enakamura@sap.ist.i.kyoto-u.ac.jp).

K. Yoshii is with Kyoto University, Kyoto, 606-8501, Japan.

The results are shown in Figure 1, where results for the connectionist quantiser [3] are omitted for clear illustration. The average (first, third quantiles) was 53.7% (43.8%, 67.3%) for the polyrhythmic data and 38.9% (28.2%, 47.3%) for the standard polyphony data. From the results we can find that methods using statistical learning tends to have better accuracies and an HMM-based model with multiple-voice structure [9] had the best accuracy for the polyrhythmic data. Example results and discussions will be given in the poster.

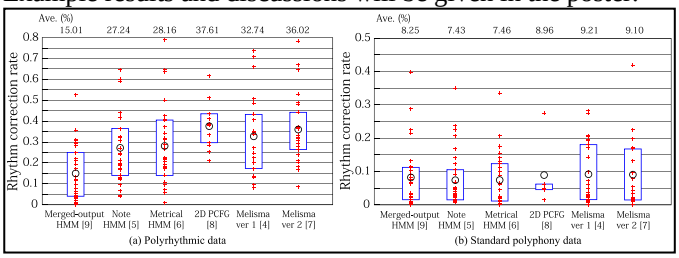


Figure 1. Rhythm correction rates (lower is better). The circle indicates the average and the blue box indicates the range from the first to third quartiles.

ACKNOWLEDGMENT

We are grateful to David Temperley, Norihiro Takamune and Henkjan Honing for providing their programming codes.

REFERENCES

[1] A. Klapuri and M. Davy (eds.), *Signal Processing Methods for Music Transcription*, Springer, 2006.

[2] E. Benetos et al., "Automatic Music Transcription: Challenges and Future Directions," *J. Intelligent Inf. Sys.*, vol. 41, no. 3, 2013, pp. 123–135.

[3] P. Desain and H. Honing, "The Quantization of Musical Time: A Connectionist Approach," *Comp. Mus. J.*, vol. 13, no. 3, 1989, pp. 56–66.

[4] D. Temperley and D. Sleator, "Modeling Meter and Harmony: A Preference-Rule Approach," *Comp. Mus. J.*, vol. 23, no. 1, 1999, pp. 10–27.

[5] H. Takeda et al., "Hidden Markov Model for Automatic Transcription of MIDI Signals," *Proc. MMSP*, 2002, pp. 428–431.

[6] C. Raphael, "Automated Rhythm Transcription," *Proc. ISMIR*, 2001, pp. 99–107.

[7] D. Temperley, "A Unified Probabilistic Model for Polyphonic Music Analysis," *J. New Music Res.*, vol. 38, no. 1, 2009, pp. 3–18.

[8] M. Tsuchiya et al., "Probabilistic Model of Two-Dimensional Rhythm Tree Structure Representation for Automatic Transcription of Polyphonic MIDI Signals," *Proc. APSIPA*, 2013, pp. 1–6.

[9] E. Nakamura et al., "Rhythm Transcription of Polyphonic MIDI Performances Based on a Merged-Output HMM for Multiple Voices," *Proc. SMC*, 2016, pp. 338–343.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Intelligent Audio Mixing Using Deep Learning

Marco Martínez[1] and Joshua D. Reiss[2]

[1]Centre for Digital Music, Queen Mary University of London, UK, m.a.martinezramirez@qmul.ac.uk
[2] Centre for Digital Music, Queen Mary University of London, UK

***Abstract*— We propose a research trajectory in the field of deep learning applied to music production systems such as mixing, mastering, sound design and sound synthesis.**

I. BACKGROUND

Multi-track audio mixing is an essential part of music production and with recent advances in machine learning techniques such as deep learning, it is of great importance to carry out research on the applications of these methods in the field of automatic mixing.

Taking into account that audio mixing is a highly cross-adaptive transformation, which, apart from artistic considerations, essentially tries to solve the problem of unmasking by manipulating different audio characteristics such as the dynamics, spatial information, timbre or pitch.

The task of automatic mixing has been researched in present years, where it was approached as an application of adaptive audio effects [1]. This was explored by the development of an intelligent system capable of performing automatic mixing [2], and an extensive understanding of the expert knowledge involved in order to develop a system of this type [3]. Machine learning has been applied to solve punctual tasks such as EQ [4] and dynamic range compression [5], but not to a whole system.

II. RESEARCH

In this research, we explore how can we train a deep neural network to perform audio mixing as a content-based transformation without using standard mixing devices (*e.g.* dynamic range compressors, equalizers, limiters, etc.), and investigating whether an intelligent system is capable of learning the intrinsic characteristics of this transformation and applies them.

In addition, we examine how such a system can perform a goal-oriented mixing task, and if we can guide it with features extracted from a mixdown. Similarly, we research how can a deep learning system use expert knowledge from the field of mixing engineering in order to improve results in an audio mixing task, and also whether we can integrate user interaction as a final fine-tuning of the mixing process.

[1]M. Martínez is a PhD researcher with the Centre for Digital Music in the School of Electronic Engineering and Computer Science at Queen Mary University of London. (author e-mail: m.a.martinezramirez@qmul.ac.uk).

[2]Dr. Joshua D. Reiss (member IEEE, AES) is a Senior Lecturer with the Centre for Digital Music in the School of Electronic Engineering and Computer Science at Queen Mary University of London. (e-mail: joshua.reiss@qmul.ac.uk)

The project will research different deep learning approaches with application to audio signals through the use of state-of-the-art training and architecture models. A very important aspect of the research will fall into this stage, since we will explore how we can understand what the system has actually learned, and if we can use this information to enhance the previous results of the system.

Through this research will also explore different ways to shape and synthesize sounds, because through training and understanding of deep learning networks, several types of test signals could be applied and new and interesting results could be discovered.

One of the main difficulties when applying deep learning to music production is the difficulty in collecting the correct data set. It is for this reason that *The Open Multitrack Testbed* will be of significant importance, with around hundreds of multi-tracks, stems and mixes, we will use this data in the process of training, developing and validation of the system. Correspondingly, we will perform an objective and subjective evaluation of the performance of the system through a series of listening experiments.

The research will focus on the post-production stage and this is where the main impact will occur, as there is a clear need for an intelligent system capable of carrying out the tedious technical tasks and also to be assist the musician or audio engineer in the creative process of audio mixing. In the same way, the research results could be of use in the field of music production, live music performance, generative music, algorithmic composition models, and intelligent music interfaces.

REFERENCES

[1] V. Verfaille, U. Zölzer, "Adaptive Digital Audio Effects (A-DAFx): A New Class of Sound Transformations.," *IEEE Transactions on Audio, Speech, and Language.,*, no. 14 (5), pp. 1817 -1831, 2006.

[2] E. Gonzalez, Advanced Automatic Mixing Tools for Music., PhD. Queen Mary University of London, 2010.

[3] P. Pestana, Automatic Mixing Systems using Adaptive Audio Effects, PhD. Universidade Católica Portuguesa, 2013.

[4] R. Reed, "A Perceptual Assistant to do Sound Equalization.," in *Intelligent User Interfaces 5th Conference*, 2000.

[5] Z. Man, B. De Man, P. Pestana, D. Black, J. Reiss, "Intelligent Multitrack Dynamic Range Compression," *Journal of the Audio Engineering Society*, vol. 63, 2015.

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Understanding the Social Character of Metadata in Music Production

Glenn McGarry[1*]

[1*]Mixed Reality Laboratory, Department of Computer Science, University of Nottingham, UK,
glenn.mcgarry@nottingham.ac.uk

***Abstract*— Musical metadata, implicated in the production of a musical object (e.g., a song), has traditionally been treated as a by-product of the production process, having little or no value beyond that, and is largely discarded. This paper explores the role of such metadata in an empirical study of a studio-based music production. It is intended to develop the findings of this and future studies into results that are tractable to the development of metadata-driven tools, to add value to music objects and support production processes.**

I. MUSICAL METADATA

Academic research has explored extraction of audio features from within the music production workflow, to enrich the metadata surrounding the process. This metadata has applications in, for example: music discovery [1]; and intelligent music production [2]. Textual metadata also has application in intelligent music production [3]; and tracking production process [4] to preserve valuable provenance information, and synthesize processual metadata to appeal to consumer interest.

Given such developments and the ubiquity of DAW (Digital Audio Workstation) software in a range of heterogeneous scenarios, the opportunities for development of new production tools and ways to create value from production metadata remains wide open. However, given the opacity of production processes [5], further work is needed to understand the social organisational role of metadata created and used in music production.

II. A MUSIC PRODUCTION STUDY

To meet these aims, an ethnographic study of a music production was undertaken, which involved observation of a recording session with a "Retro Rock" covers band. The engineer on the session created most of the metadata in the form of textual labeling to support his working methods and processes. Labels were applied to recording hardware and software at key positions visible from a working position in the studio control room to indicate the signal path of an audio source. This enabled tracking of progress through 'set up' activities, and aided navigation of recording equipment controls during the task of monitoring (listening for audio quality and watching level meters), and balancing instrument levels during the recording. Labeling also substituted for note taking of vital multi-track; configuration; instrumentation; song structure; and editorial information to be used later in post-production editing and mixing process.

The findings of this and one other study [6] have been taken up by project developers (see acknowledgment), who propose to develop a demonstrator to automatically label audio using feature extraction, to support the production workflow [7].

III. CONCLUSION

The analysis of the fieldwork has unpacked much about the social life of music production, the actors involved in making music; the activities and collaborations they engage in; the resources (metadata) they exploit; and how music is crafted in practice; to drive design thinking around new value propositions exploiting metadata.

This study however, represents only one small part of a large problem area and further studies are in progress to understand different music making methods and scenarios, and other stages of the music production process.

ACKNOWLEDGMENT

I would like to acknowledge the grant Fusing Semantic and Audio Technologies for Intelligent Music Production and Consumption - EP/L019981/1

REFERENCES

[1] Kolozali, S. (2014). A framework for automatic ontology generation based on semantic audio analysis, 1–10.

[2] Fazekas, G., & Sandler, M. B. (2007). Intelligent Editing of Studio Recordings with the help of Automatic Music Structure Extraction. *Proceedings of the 122nd AES Convention*.

[3] Wilmering, T., Fazekas, G., & Sandler, M. B. (2012). High-Level Semantic Metadata for the Control of Multitrack Adaptive Digital Audio Effects. In *Proceedings of the 133rd AES Convention*.

[4] Barkati, K., Bonardi, A., Vincent, A., & Rousseaux, F. (2012). GAMELAN: A Knowledge Management Approach for Digital Audio Production Workflows. *Proceedings of Artificial Intelligence for Knowledge Management Workshop of ECAI 2012*, 3–8.

[5] Paul Thompson and Brett Lashua. 2014. "Getting it on Record Issues and Strategies for Ethnographic Practice in Recording Studios." *Journal of Contemporary Ethnography (2014)*.

[6] McGarry, G., Tolmie, P., Benford, S., Greenhalgh, C., & Chamberlain, A. (2017). "They're all going out to something weird": Workflow, Legacy and Metadata in the Music Production Process. In *Proceedings of the 2017 ACM Conference on Computer-Supported Cooperative Work and Social Computing*. Portland, USA.

[7] http://www.semanticaudio.ac.uk/blog/ethnographic-studies-of-studio-based-music-production/)

DMRN+11: DIGITAL MUSIC RESEARCH NETWORK
ONE-DAY WORKSHOP 2016

QUEEN MARY UNIVERSITY OF LONDON
TUE 20 DECEMBER 2016

Automatic Transcription of Vocal Quartets

Rodrigo Schramm[1*] and Emmanouil Benetos[2]

[1*]UFRGS, Brazil and C4DM/Queen Mary University, United Kingdom, r.schramm@qmul.ac.uk
[2] C4DM/Queen Mary University, United Kingdom, emmanouil.benetos@qmul.ac.uk

***Abstract*—This work presents a probabilistic latent component analysis (PLCA) method applied to automatic music transcription of *a cappella* performances of vocal quartets. A variable-Q transform (VQT) representation of the audio spectrogram is factorised with the help of a 6-dimensional tensor. Preliminary experiments have shown promising music transcription results when applied to audio recordings of Bach Chorales and Barbershop music.**

I. INTRODUCTION

Automatic music transcription is a process that converts audio signals into a symbolic representation (such as a music score) and can further be used to support applications in music informatics, musicology, interactive music systems, and automatic music assessment. Despite recent advances in the field, automatic music transcription of vocal quartets has not yet been explored, and it is still considered an open problem. The use of spectrogram factorization algorithms, such as non-negative matrix factorization (NMF) and probabilistic latent component analysis (PLCA) have been extensively used [1,2,3,4] in the task of audio source separation and multi-pitch estimation in the last decade. The main idea of these approaches is to explain an input time-frequency representation (spectrogram) by a linear combination of non-negative factors, consisting mainly of spectrum atoms and note activations.

II. MODEL

In this approach, we implement a PLCA-based model following the idea of [5] which uses fixed dictionary templates, representing the log-spectrogram (time-frequency) through a variable-Q transform. In our approach, the primary purpose is to explore the factorization of the log-spectrogram into components that have a close connection with singing characteristics as voice type (soprano, contralto, tenor, baritone and bass) and the vocalization of vowels. Thus, we formulate the templates into a 6-dimensional tensor, representing pitch, voice type, vowel, singer source, log-frequency index, and shifting across log-frequency (5 bins per semitone). As a similar approach to [3], the singer source parameter constrains the search space into mixture-of-subspaces, clustering a large range of singers (instruments) into a small number of categories. This configuration allows us exploring temporal coherence and sparsity among these characteristics and, consequently, improving the final accuracy of the multi-pitch estimation.

III. EXPERIMENTS

Experiments were conducted to evaluate the proposed model by using recordings of Bach Chorales and Barbershop music with performances of vocal quartets. This audio dataset was built with recordings available in http://pgmusic.com. Our experiments include mixes with two, three and four voices. Preliminary results have shown good estimates of the multiple fundamental frequencies (multiple pitches) from these polyphonic recordings when applying a fixed thresholding step on the multi-pitch activation output. However, overtones are yet responsible for the false detection of pitches in higher octaves. As future work, we plan to implement a post-processing procedure based on Factorial Hidden Markov Models [6], avoiding the influence of overtones and producing a more consistent music transcription and voice separation output.

*Research supported by the Royal Academy of Engineering under NRCP1617/5/46. EB is supported by a RAEng Research Fellowship (grant no. RF/128).

Dr. R. Schramm is a Lecturer in the Music Department, UFRGS/Brazil, and an Academic Visitor in the C4DM, Queen Mary University of London.

Dr. E. Benetos is a Research Fellow and Lecturer in the C4DM, Queen Mary University of London.

REFERENCES

[1] Emmanouil Benetos and Simon Dixon. A shift- invariant latent variable model for automatic music transcription. *Computer Music Journal*,36(4):81–94, 2012.

[2] B. Fuentes, R. Badeau, and G. Richard. Controlling the convergence rate to help parameter estimation in a plca-based model. In *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, pages 626–630, Sept 2014.

[3] G. Grindlay and D. P. W. Ellis. Transcribing multi-instrument polyphonic music with hierarchical eigeninstruments. *IEEE Journal of Selected Topics in Signal Processing*, 5(6):1159–1169, Oct 2011.

[4] Hirokazu Kameoka, Masahiro Nakano, Kazuki Ochiai, Yutaka Imoto, Kunio Kashino, and Shigeki Sagayama. Constrained and regularized variants of non-negative matrix factorization incorporating music-specific con- straints. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2012, Kyoto, Japan, March 25-30, 2012*, pages 5365– 5368, 2012.

[5] Emmanouil Benetos and Tillman Weyde. An efficient temporally-constrained probabilistic model for multiple-instrument music transcription. In *Proceedings of the 16th International Society for Music In- formation Retrieval Conference, ISMIR 2015, Málaga, Spain, October 26-30, 2015*, pages 701–707, 2015.

[6] M. Wohlmayr and F. Pernkopf, "Model-Based Multiple Pitch Tracking Using Factorial HMMs: Model Adaptation and Inference," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 8, pp. 1742-1754, Aug. 2013.