CogMIR™

Cognitively Based Music Informatics Research

Program Schedule
August 27, 2012

Moderator: Paolo Ammirante, Ryerson University
Invited Speakers:

- Keynote 1: Michael Casey, Dartmouth College
- Keynote 2: George Tzanetakis, University of Victoria
- Matthew Woolhouse, McMaster University

Other Speakers:

- Naresh Vempala, Ryerson University
- Frank Russo, Ryerson University
- David Beckford / Glen Kappel, *wave*DNA
- Steven Livingstone, Ryerson University

Schedule:
09:15 am-09:50 am – Coffee/Snacks

09:50 am-10.00 am – Introduction

10:00 am-11:00 am – Talk 1: Michael Casey (Keynote)
Mapping Mental Representations of Music with fMRI and MIR

11:00 am-12:00 pm – Talk 2: George Tzanetakis (Keynote)
Blending the Physical and the Virtual: Two Case Studies of Embodied Musicianship

12:00 pm-12:30 pm – Talk 3: Naresh Vempala
Applying Computational Modeling Approaches to Music Emotion Perception

12:40 pm-01:35 pm – Lunch / Posters

01:35 pm-03:00 pm – Student Posters and Talks 4-10

03:00 pm-03:30 pm – Talk 11: Frank Russo
Building Bridges Between Music Cognition and Music Informatics

03:30 pm-03:45 pm – Coffee

03:45 pm-04:15 pm – Talk 12: David Beckford/Glen Kappel
Crossover Toolkits for Music Production and MIR: Navigating Rhythmic Similarity in Liquid Loops

04:15 pm-04:45 pm – Talk 13: Matthew Woolhouse (Invited Talk)
Drawing Inferences from Music-download Metadata

04:45 pm-05:05 pm – Talk 14: Steven Livingstone
Changing Musical Emotion with a Computational Rule System

05:05 pm-05:15 pm – Concluding remarks

Talk 1 (Keynote): Michael Casey, Dartmouth College

Mapping Mental Representations of Music with fMRI and MIR

In this talk, we show that representational spaces used in everyday music listening are revealed by combining music information retrieval (MIR) and functional magnetic resonance imaging (fMRI) via multivariate pattern analysis (MVPA). First, we present evidence that neural blood oxygenation level dependent (BOLD) activation patterns evoked by listening to rich music recordings, measured by fMRI, are predictive of musical style for novel stimuli. A multi-class linear support vector machine classifier predicted musical style, for novel music in five categories, from bi-lateral superior temporal sulcus (STS) with 55% accuracy vs. 20% baseline. A follow-up behavioral experiment showed the classifier confusions to be consistent with human perception. Accuracy dropped significantly when classifying from primary auditory cortex (PAC) alone, suggesting that superior temporal cortex encodes additional important high-level information about music. To probe the type of music information encoded in PAC and STS neural populations we used representational similarity analysis (RSA) and predictive multivariate regression between MIR and fMRI features. Our results show which features, among a large collection of common MIR audio and schematic features, were significantly predictive of BOLD activation patterns, thereby allowing interpretation of the musical meaning embedded in each neural population. We argue that joint fMRI MIR multivariate pattern analysis reveals the representational structure of the brain listening to music. Our final study shows that we are able to predict crowd-sourced musical tags from human brain data using MVPA, thereby revealing semantic representational codes. Such analyses could lead to significant cognitively-informed improvements for music information processing systems.

Talk 2 (Keynote): George Tzanetakis, University of Victoria

Blending the Physical and the Virtual: Two Case Studies of Embodied Musicianship

Researchers in embodied cognition posit that our cognition is influenced by our sensory perception, motor control and situatedness. In the field of robotics this idea has been used to argue that true artificial intelligence is not feasible unless machines have sensory and motor skills and are able to interact with the world through a body. Robert Rowe in this book Machine Musicianship explores how computer systems can benefit from a systematic foundation of musical knowledge. Music information retrieval research has mostly focused on how to extract information at various levels of detail from large collections of music in digital representations including audio signals. In both of these cases the focus is more on issues of cognition and representation rather than any notion of a body. In this talk I will discuss why I believe the concepts in embodied cognition are important in building better systems for music making involving computers as well as understanding the process of music creation. I will support my discussion through two concrete examples of systems that utilize concepts from music information retrieval and in which there is a clear notion of embodiment and a blending of the physical and virtual world:

1) Physical modelling synthesis refers to methods that synthetically generate musical instrument sounds by using a set of equations and algorithms that simulate the physics of sound production. They provide realistic sounds with controls that are physically meaningful. However the control of physical modelling algorithms is challenging. Using machine learning techniques we show how a virtual violinist can "learn" to bow in a similar way to a beginning violin student.

2) Robotic percussion instruments are typically simple output devices that are essentially deaf. I will describe how they can be imbued with the ability to listen to themselves and other musicians in the context of sound checking as well as live music improvisation involving both machines and humans.

Talk 3: Naresh Vempala, Ryerson University

Applying Computational Modeling Approaches to Music Emotion Perception

In this talk, we describe our implementation of two feedforward neural networks for predicting mean valence/arousal ratings of participants, for musical excerpts based on (a) audio features, and (b) physiological changes. Previous studies have shown that audio features may be used to predict emotion judgments of listeners. Previous studies have also shown that physiological changes occur during music listening that correspond to felt emotion. However, the nature of this relationship is somewhat unclear. Our goal was to better understand this relationship between music emotion judgments, audio features, and physiological responses. Thirteen audio features were extracted from 12 classical music excerpts (3 from each emotion quadrant). Valence/arousal ratings, and five different physiological responses were collected from 45 participants. Both networks were trained on eight of the 12 excerpts and tested on the remaining four. The networks predicted values that closely matched mean participant ratings of valence and arousal. Our study indicates that neural networks can be trained to identify statistical consistencies across audio features as well as physiological changes to predict valence/arousal values. We discuss the implications of these results for music emotion judgment.

(For student poster talks 4-10, see page 5)

Talk 11: Frank Russo, Ryerson University

Building Bridges Between Music Cognition and Music Informatics

In this talk, I will discuss cognitively based music informatics research (cogMIR) as a new bridge that is connecting two increasingly prominent subfields of music research: Music Cognition (MC) and Music Informatics (MI). While the two subfields often deal with related problems, the objectives and methods used to address the problems tend to be orthogonal. For example, both subfields may be interested in developing computational models that quantify the similarity between fragments of music. However, MI seeks to develop computationally efficient models, whereas MC seeks to develop models that bear resemblance to human judgment. The talk will address whether cogMIR has the potential to seed a paradigm shift, how music researchers can adopt a cogMIR approach, and provide a survey of the low-lying fruit.

Talk 12: David Beckford, Lead Inventor and CSO, WaveDNA
Glen Kappel, Lead Researcher and CIO, WaveDNA

Crossover Toolkits for Music Production and MIR: Navigating Rhythmic Similarity in Liquid Loops

Members of WaveDNA will report on progress achieved since CogMIR 2011 concerning crossover toolkits to aid in both creative music production as well as music informatics research. WaveDNA's research team will provide a brief recap and overview of their proprietary symbolic music representation system, highlighting the components of their bar-length "Music Molecule"™.

Two new system-based "rhythmic similarity maps" will be introduced that group and organize the molecular components used to either create or analyze drum and percussion loops. A current implementation of these rhythmic navigation maps within a new software drum and percussion loop generator application (the Liquid Loops Beatweaver Rhythm Synthesizer) will also be briefly demonstrated.

Current research and development on future extensions of these rhythmic similarity maps will also be disclosed and demonstrated in prototype with the aid of student developers from the WaveDNA team. These extensions (alluded to in vague terms at CogMIR 2011) begin to deliver on the possibility of navigating musical similarity spaces by way of cognitively-based attributes, and have been developed based upon experiments conducted in collaboration with Ryerson's SMART Lab. Short demonstrations of ways in which such 'Cognitive Music Informatic tools' can be used to either analyze existing rhythms or create new drum and percussion loops within a commercial software application (future versions of Liquid Loops) will be shown.

Talk 13 (Invited Talk): Matthew Woolhouse, McMaster University

Drawing Inferences from Music-download Metadata

This presentation will report findings from a 2010 study in which 1.8 million anonymised music downloads from 13 countries, supplied by the Nokia Corporation, were analyzed from sociocultural perspectives. The data were generated by Nokia's online music stores, which allow users to purchase and download music directly onto a mobile device or home computer. The research explored factors such as the level of musical diversity and individual adventurousness within each country, and download patterns and genre changes across 24 hours. Some of these findings were then correlated with demographic measures such as language acquisition and migration. In addition, countries' musical genre preferences were matched with music preference-personality measures (North, 2010; Rentfrow & Gosling, 2003) in order to derive a profile of "national characteristics" for each country. The research revealed numerous intriguing differences and similarities between countries from the Americas, Europe, the Middle East and Asia.

Talk 14: Steven Livingstone, Ryerson University

Changing Musical Emotion with a Computational Rule System

Musicians communicate emotional intentions through the control of acoustic features such as pitch, loudness, and articulation. The extent to which musical emotion can be controlled by software through the manipulation of these features has not been fully examined. In this talk I will present CMERS, a Computational Music Emotion Rule System for the real-time control of musical emotion. The effectiveness of CMERS was assessed with perceptual testing. Listeners continuously rated the arousal and valence of musical works that were modified by CMERS to express a range of different emotions. Significant shifts in arousal and valence were recorded for all manipulations. Listeners correctly selected the intended system emotion with 78% accuracy. This study demonstrates the effectiveness of using computational methods to study musical emotions.

Student Poster Talks

Talk 4: *FeatureFinder*: *A Powerful Signal Processing Tool*
http://www.featurefinder.ca/

Alex J. Andrews*, Gabe Nespoli*, Frank A. Russo*
SMART Lab, Department of Psychology, Ryerson University

Physiological features can provide a useful supplement to audio and tags in music informatics research. However, large-scale processing of physiological features is extremely time consuming and prone to human error. *FeatureFinder* is a freely-available software package that lets you organize, filter, and analyze physiological signals (e.g., EMG, HR, GSR) in a manner that is more extensible than commercially available software. The user can enter a large number of data files in one step and then instantly preview effects of different filters and feature extraction methods. Processing is executed via a graphical-user-interface that allows for the development of novel filters, features, and processing streams. Features are then extracted to a text file for subsequent statistical analyses and/or classification. This talk is intended to showcase *FeatureFinder*'s functionality and to expand its growing community of users.

Talk 5: *The Impact of Hearing Loss on Music Cognition*

David Fourney*
***Department of Mechanical & Industrial Engineering**
Ryerson University

Research relating to music cognition tends to assume the human listener has perfect stereo hearing yet this is generally not the case. Hearing loss is the third most chronic disability among older adults and the fastest growing hidden disability in North America. Among Canadians aged 45 to 87, 46% have a hearing loss. An increasingly aging global population means the number of people who are Deaf, Deafened, or Hard of Hearing (D/HOH) is expected to grow. This talk will discuss the results of a series of interviews with D/HOH people exploring their music experiences. Eight people (6 f, 2 m) aged 18 – 65+ who all identified themselves as HOH participated in one-on-one interviews. All but three were born

hearing and acquired their hearing loss later in life. Major themes identified included: participants all enjoyed music prior to their hearing loss and actively try to access it afterward, a strong reliance on specific strategies (e.g., past memory of music) to help enjoy and understand music, hearing loss changes what music sounds like (e.g., participants reported "gaps", "static", "flatness"). New research questions, such as whether a person's hearing loss influences their specific music choices, were also identified. A better understanding of the music cognition of D/HOH people would assist in the development of technologies to support D/HOH music consumers.

Talk 6: ***Cognitively-based Rhythmic Informatics in an Information Age Drum Machine***

Matthew Giamou*, Natasha Dalal*, Naresh N. Vempala^, Frank A. Russo^
***University of Toronto, WaveDNA**
^SMART Lab, Ryerson University

This talk will involve a presentation of two new analytic tools built upon the WaveDNA Music Molecule framework that focus on methods of organizing the rhythmic components featured in their recently introduced Rhythm Synthesizer and Beat Generator, Liquid Loops. Student developers from WaveDNA will introduce a new widget that features an organization of bar-length "BarForm" components according to the results of experiments conducted in collaboration with the SMART Lab team. This widget acts as a graphing tool that plots all the possible 'activation patterns' for rhythmic patterns in 4/4-time, in a 2-dimensional space depicting the relative 'stability' and 'activity' of the various patterns as judged by the participants in a cognitive study. This 'cognitive-overlay' widget works in conjunction with added functionality, also designed and implemented by the student presenters, that operates upon the system-based navigation wheels that currently feature in Liquid Loops. Beatseeker, a graphical representation of instrument-dependent Barform co-occurrence, uses color transparency to visualize the relative frequency of individual Barforms on the activity wheel. Selectively combining these two maps in various ways against an existing corpus of musical examples can allow a user to highlight the usage of particular Barforms in particular contexts on the BarForm navigation wheel. The graphing tool is extensible to any 2 dimensional dataset. These tools are examined and demonstrated as both analytical and compositional aids.

Talk 7: ***Audio Features as Predictors of Tiredness and Anxiety in Music: A Study Across Two Genres***

Lisa Liskovoi*, Naresh N. Vempala*, Colleen E. Carney*, Naomi Koerner*, Frank A. Russo*
Department of Psychology, Ryerson University

This study attempted to create a linear regression model capable of identifying tired and anxious pieces using audio features. Although tiredness and anxiousness are both negatively valenced, these emotions are qualitatively different in their cognitive and behavioural effects. The model was developed using a dataset of 40 excerpts from classical and soundtrack genres. Twelve features, extracted using MIRToolbox, were used as predictors of subjective evaluations. We performed three sets of regressions; all tracks, classical only, and soundtrack only. Different feature-combinations served as best predictors of emotions across genres. Combining classical and soundtrack pieces, spread and brightness were predictive of tiredness; RMS, brightness and centroid predicted anxiousness. For soundtrack pieces, increases in tiredness were best predicted by increase in spectral spread and decrease

in pulseclarity while anxiousness was best predicted by increase in RMS and increase in pulseclarity. For classical pieces increases in tiredness were predicted by decrease in RMS, and tempo, and decrease in brightness; increases in anxiousness were predicted by increase in RMS and zerocross. These models were tested by calculating predicted scores for 12 additional pieces. Mean difference between predicted and actual ratings revealed that the combined genre model was the most effective at predicting tiredness, while genre specific models were better at predicting anxiousness.

Talk 8: ***Predicting Crowdsourced Musical Tags from Brain Activity***

Jessica Thompson*, Michael Casey*, Lorenzo Torresani^
***Bregman Music and Auditory Research Studio, Department of Music**
^Department of Computer Science
Dartmouth College

Tags are a growing source of music information that can shed light on high-level judgments of music. Here we investigated the encoding of distinct musical concepts in the human brain via their relationship to tags. Blood oxygenation level dependent signals were recorded via functional magnetic resonance imaging while 15 subjects listened to 25 clips of music from 5 different musical styles. A dictionary of the 100 most popular tags for the 25 stimuli was selected using the last.fm API, which gives access to tags from millions of users. Last.fm tags were successfully predicted from brain activity in superior temporal cortex using a k-nearest neighbors classifer. Precision@k was evaluated for each subject and for k={1, 5, 10} using leave-one-out cross validation. Precision values were compared to a baseline precision calculated by assigning tags at random. Tag prediction precision was significantly greater than random for all subjects and all k, performing up to 28% better than the random baseline. We conclude that high-level information used to label and categorize music is represented in distributed population codes in superior temporal cortex. We argue that a better understanding of the underlying neural representation of tags can help to inform automatic media annotation systems.

Talk 9: ***Online Learning of Emotion Rating Changes using Musical Features as Experts***

Finn Upham*
***Steinhardt School of Culture, Education, and Human Development**
New York University

Online learning with musical feature experts allows for dynamic serial prediction of emotional response series without interference from collinearity or the presumptions of static relationships between musical cues and listeners' responses. This paper presents the basics of online learning, the process of adapting continuous ratings of emotion and musical features to suit the online paradigm, and the results of modelling individual listening responses. Initial results suggest that participants change in sensitivity to musical features during listening, and while prediction power is very poor for these first applications of online learning algorithms, they present novel descriptors of individual responses for future classification.

Talk 10: ***The Effect of Rhythmic Distortion on Melody Recognition***

David M. Weigl*, Catherine Guastavino*, Daniel J. Levitin^
***School of Information Studies**
^Dept. of Psychology
McGill University
Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT)

Melodies consist of intertwined musical facets. Investigations of their relative contributions to melody identification have applications in the development of experiential relevance and similarity measures for MIR. Existing studies typically employ equitonal and isochronic distortions of familiar melodies in melody recognition tasks. Results have indicated significantly lower recognition rates under equitonal relative to isochronic conditions, implying a diminished role of rhythm compared to pitch. These studies share the assumption that imposing isochrony nullifies rhythmic information. However, the familiar melodies typically exhibit very simple rhythms, and the degree of distortion is not quantified. The present study employs a set of more complex rhythmic distortions, measuring the degree of distortion using the chronotonic distance, to address this issue. Distorted versions of familiar melodies are presented to two groups of participants in the following conditions: shuffled (durations randomly reassigned among the notes of the melody); randomized (all notes assigned random durations); stretched (melody uniformly slowed to match the duration of the randomized condition); isochronic (all notes assigned a constant duration), and undistorted. Results indicate significantly lower recognition rates in the randomized condition, compared to the other conditions. This suggests that the isochronic distortion used in previous studies insufficiently nullifies rhythmic information. Our findings further highlight the contribution of rhythmic information to melody recognition.