CogMIR™

August 12, 2016

Columbia University, Northwest Corner Building (NWC), Room 501

550 West 120th Street, New York, NY

Schedule:

08:30 am-09:00 am	Registration
09:00 am-09.10 am	Welcome (Naresh Vempala/Frank Russo)
09:10 am-09:55 am	Keynote – **Daniel Müllensiefen:** *Individual differences and cognitive music information retrieval*
10:00 am-11:00 am	**Kaustuv Kanti Ganguli & Preeti Rao:** *Exploring melodic similarity in Hindustani classical music through the synthetic manipulation of raga phrases*
	Bastiaan van der Weij, Marcus Pearce, Henkjan Honing: *A probabilistic model of meter perception*
	Jordan Smith & Elaine Chew: *Validating a technique for post-hoc estimation of a listener's focus in music structure analysis*
	Anders Elowsson & Anders Friberg: *A neural network model of performed dynamics*
11:00 am-11:15 am	Break
11:15 am-12:15 pm	**Speed poster talks** (see page 2)
12:15 pm-01:30 pm	Lunch/Poster session
01:30 pm-02:15 pm	Keynote – **Juan Pablo Bello:** *Some thoughts on the how, what and why of music informatics research*
02:20 pm-03:20 pm	**John Ashley Burgoyne & Jan Van Balen:** *Predicting long-term musical memorability with the CATCHY toolbox*
	Naresh Vempala, Michael Barone, Matthew Woolhouse: *Identification of listener genre preference using the Echo Nest API features*
	Ji Chul Kim & Edward Large: *Multiple F0 estimation by Gradient Frequency Neural Networks*
	Sebastian Stober: *Learning about music cognition by asking music information retrieval questions*
03:20 pm-03:35 pm	Break
03:35 pm-04:20 pm	**Johanna Devaney:** *Comparing human and computer performance on a music similarity task*
	Gen Hori et al.: *Musical chord progression and alpha band desynchronization of EEG*
	Carson Miller Rigoli, Sarah Creel, Petr Janata: *Neural resonance and beat tracking in complex meters*
04:25 pm-04:45 pm	Announcement of Student Prizes/Closing Remarks

Partners

Smart Lab

ISMIR

Ryerson University

Columbia University

Sponsors

SHAZAM

NSERC CRSNG

Next

Speed poster talks:

1. **Matevz Pesek, Ales Leonardis, Matija Marolt**
   Pattern discovery and music similarity with compositional hierarchical model
2. **Trevor de Clercq**
   Tempo versus harmonic and melodic pacing in a corpus of rock music
3. **Iris Yuping Ren, David Temperley, Zhiyao Duan**
   Blue notes in rock: An exploratory study
4. **David Heise, Jason Noble, Rhimmon Simchy-Gross**
   On machine perception of sound mass as a step toward computational auditory scene analysis
5. **Anders Friberg, Ragnar Schon, Anders Elowsson, Khayun Choi, Stephen Downie**
   Cross-cultural variations among rating of perceptual features in Korean popular music
6. **Scott A. Miles, David S. Rosen, Norberto Grzywacz**
   The role of surprise in music preference
7. **Blair Kaneshiro, Tom Collins, Anthony M. Norcia, Jonathan Berger**
   Using representational similarity analysis to study perception of tonal categories
8. **Andrew Lambert, Tillman Weyde, Newton Armstrong**
   Adaptivity in oscillator-based pulse and metre perception

Keynote Addresses:

Daniel Müllensiefen, Goldsmiths, University of London

Individual differences and cognitive music information retrieval

Music is a domain that is notorious for the diversity of judgments and opinions that a single item (i.e. piece of music) can create. People differ in their musical tastes, their musical knowledge and skills, in their language to describe and label music and even more fundamentally in how they perceive and categorize salient events in a musical piece. But if people can rightfully disagree on their perceptions and evaluations of music, then this creates fundamental problems for algorithms and cognitive systems that try to understand music-related behavior by a mechanism that does not account for individual differences between listeners, leading to datasets with high und seemingly unstructured variability. As potential causes for highly variable datasets I discuss scenarios where listener responses have low reliability or validity and where several solutions to a given music cognition task exist. I then suggest some remedies to account for individual differences effects including the use of large samples and repeated measurements, the use of principles for test construction from modern psychometrics, and the inclusion of standardized individual differences measures into cognitive and computational models. Finally we'll sketch out how joining the perspectives of cognitive and individual differences psychology can be leveraged to create more comprehensive and more robust models in music information retrieval.

Juan Pablo Bello, New York University

Some thoughts on the how, what and why of music informatics research

The framework of music informatics research (MIR) can be thought of as a closed loop of data collection, algorithmic development and benchmarking. Much of what we do is heavily focused on the algorithmic aspects, or how to optimally combine various techniques from e.g., signal processing, data mining, and machine learning, to solve a variety of problems, from auto-tagging to automatic transcription, that captivate the interest of our community. We are very good at this, and in this talk I will describe some of the know-how that we have collectively accumulated over the years. On the other hand, I would argue that we are less proficient at clearly defining the "what" and "why" behind our work, that data collection and benchmarking have received far less attention and are often treated as afterthoughts, and that we sometimes tend to rely on widespread and limiting assumptions about music that affect the validity and usability of our research. On this, there is much that we can learn from music cognition research, particularly with regards to the adoption of methods and practices that fully embrace the complexity and variability of human responses to music, while still clearly delineating the scope of the solutions or analyses being proposed.

Abstracts:

(listed in alphabetical order by first author's last name)

Predicting long-term musical memorability with the CATCHY toolbox

John Ashley Burgoyne* & Jan Van Balen**

***University of Amsterdam**

****Utrecht University**

At ISMIR 2015, Van Balen et al. presented a set of audio features for use in cognitive experiments, now available as the CATCHY toolbox (https://github.com/jvbalen/catchy). The CATCHY features have been used to study the musical characteristics most associated with long-term recall of musical imagery, including a model for predicting the relative memorability of different structural sections of popular Western music. As an example of how these features and this model might be applied more broadly, we apply them to a very different corpus: the 43 entries to the 2016 Eurovision Song Contest. We segmented the audio for these songs using the Echo Nest segmenter, computed CATCHY features, and applied our memorability model to each segment. Although the model was never intended for hit-song science, it had some success as a predictor of Eurovision contest performance, our top 10 capturing 3 out of the actual top 10 and 4 out of the top 10 from televoting, compared to a chance rate of 2.3. The results are more enlightening for insight into how the model works in practice, highlighting in particular the outsized importance of the 'vocal prominence' factor. We will discuss the broader implications of using the toolbox and model predictions for cognitive MIR.

Tempo versus harmonic and melodic pacing in a corpus of Rock music

Trevor de Clercq

Middle Tennessee State University

As the tempo of a song increases, the duration of harmonic and melodic events will decrease proportionally in terms of absolute time, e.g., a chord lasting four seconds will last only two seconds if the tempo doubles. Examining a large corpus of rock music, however, I find that this principle does not hold true when comparing

tempo to harmonic and melodic pacing between songs. Specifically, median chord lengths and median melodic note durations per song—as measured in seconds—generally remain constant across the entire range of tempos. Relative harmonic and melodic lengths thus tend to increase as tempo increases, e.g., melodies in songs at slower tempos tend to use eighth notes, while those at faster tempos tend to use quarter notes. I argue, therefore, that absolute time shapes not only the perceptual limits for meter and maximal pulse salience (London 2012) but also cognitive preferences for rates of harmonic and melodic motion. Most importantly, these results help explain why different listeners often entrain to different metric levels in rock music (Moelants and McKinney 2004, Levy 2011), since the speed of the drum pattern may be considered distinct from the rate at which harmonic and melodic content is disbursed.

Comparing human and computer performance on a music similarity task

Johanna C. Devaney
School of Music, The Ohio State University

The ability to parse the music surface into structurally significant and non-significant notes facilitates a hierarchical understanding of musical works and can reveal structural similarities between different musical pieces. This paper considers the utility of incorporating human domain knowledge of music analysis into a computational algorithm for estimating music similarity. This utility is measured using a six-alternative forced choice matching experiment in both the symbolic and audio domains using six sets of Beethoven themes and variations. In these pieces, the harmonic progressions remain stable in each variation while melodic characteristics and textures vary. A computer experiment compared the performance of representations using notes versus chord tones in the symbolic domain and chroma versus "chord-tone chroma" in the audio domain on the matching task. As expected, pruning by chord tone and "chord-tone chroma" improved the performance of the computational algorithms on the matching task, showing the positive effect of including domain knowledge. When the results of a pilot experiment, which consisted of 8 human participants, were analyzed against the computer algorithms, the computer performed better overall but showed greater variability in both the symbolic and audio tasks. Data analysis is currently ongoing for a further 22 participants to verify these findings.

A neural network model of performed dynamics

Anders Elowsson & Anders Friberg
KTH Royal Institute of Technology, Stockholm, Sweden

By varying the dynamics in a performance (e.g. p, ff), the performer can shape the expressive character of the music. Specifically, performed dynamics have been shown to be an important parameter for the emotional expression. In this study we present a model for identifying performed dynamics in musical audio. Ground truths were collected by asking listeners to rate the performed dynamics (soft/loud) on a semi-continuous scale. Ratings were collected from about 20 listeners for two datasets, totaling 210 musical audio excerpts. The musical excerpts were filtered with signal processing methods that captured primarily spectral fluctuations at relevant points in time, such as tone starts. The listener ratings, averaged over subjects, were then used as targets to train an ensemble of neural networks. We used 500 neural networks in the ensemble, each trained with a subsample of the computed features. The model was able to capture performed dynamics accurately ($R2 = 0.84$), a result that is close to the upper bound given the estimated uncertainty of the ground truth data. The results are well above those of individual human listeners of the experiment, and on par with the combined performance of 6 listeners.

Cross-cultural variations among ratings of perceptual features in Korean popular music

Anders Friberg*, Ragnar Schön[+], Anders Elowsson*, Khayun Choi^, Stephen Downie^
***KTH Royal Institute of Technology, Stockholm, Sweden**
[+]Swedish Radio, Stockholm, Sweden.
^University of Illinois, Champaign

In previous studies we have shown that using the concept of perceptual features as an intermediate representation can be a relevant strategy for modeling higher-level semantic descriptions, such as emotional expression. In this study we focused on possible cross-cultural differences regarding ratings of perceptual features. To investigate this, we asked both Chinese and Swedish listeners to rate a set of K-Pop samples using a web-based questionnaire. The music samples were selected from a larger set that was previously rated in terms of different emotion labels. The selection procedure was carefully designed to maximize the variation of both emotion and genre in the resulting subset. The listeners rated eight perceptual features: dissonance, speed, rhythmic complexity, rhythmic clarity, articulation, harmonic complexity, modality and pitch. The results indicated a small but significant difference of the mean values for speed and rhythmic complexity while the other features were non-significant. The overall consistency of the ratings were lower than in previous studies and differed between the two groups. This also influenced the prediction of the emotion dimensions activity and valence. We discuss the possible reasons for the lower consistency and the cultural differences in relation to the interface, musical background, and number of subjects.

Exploring melodic similarity in Hindustani classical music through the synthetic manipulation of raga phrases

Kaustuv Kanti Ganguli & Preeti Rao
Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India

A raga performance in North Indian classical music builds upon a melodic framework where the characteristic phrases of the raga appear repeatedly and with considerable creative variation while strongly retaining their identity. It is of interest for both, music retrieval and pedagogy, to understand better the space of "allowed" variations of the melodic shape corresponding to a raga phrase. We present a study of melodic shapes corresponding to a selected raga phrase extracted from performances by eminent vocal artists, which leads us to two main dimensions of variation, namely, the precise intonation of steady notes and temporal extent of a passing note. Several synthetic but musically consistent versions of the phrase are generated from the canonical form and presented to musicians in a rating task related to the raga identifiability of the stimulus. We observe that the ratings are nearly categorical with differences above a threshold corresponding to a gross change in perception from the given raga to a different raga. We intend to extend the study to brain waves and report the measurement of the P300 ERP in an oddball paradigm based perception test to strengthen the results of the behavioral experiment.

On machine perception of sound mass as a step toward computational auditory scene analysis

David Heise*^, Jason Noble[+]^, Rhimmon Simchy-Gross[#]^
***Department of Computer Science, Technology & Mathematics, Lincoln University**
[+]Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Schulich School of Music, McGill University
[#]Music Cognition Lab, University of Arkansas
^CRoMA (Computational Research on Music & Audio) Team of Interdisciplinary

Collaborators (CRoMA-TIC), Lincoln University

Sound mass has been a subject of interest in musical composition since at least the 1950s and has more recently been studied perceptually. Douglas, Noble, and McAdams (2016) suggest that sound mass is "a type of auditory grouping that retains an impression of multiplicity even as it is perceived as a perceptual unit." This qualitative definition distinguishes a sound mass from other discrete auditory events (e.g., single musical notes), but we may be able to use the quantitative characteristics of sound masses to detect perceptually relevant events, providing an intriguing approach to computational auditory scene analysis (CASA). A system could rate the likelihood that a particular segment of audio (in time-frequency space) belongs to a sound mass source, which may be an aggregation of multiple individual sources. We are developing a framework to implement exactly this, beginning with a functioning system that considers attack rate, register, and pitch density as inputs within temporal windows of an audio signal to compute and output a novel measure of sound mass density (SMD). We will demonstrate how the outputs of this system correlate with the experimental findings of Douglas et al. with regard to sound mass perception in Ligeti's Continuum.

Musical chord progression and alpha band desynchronization of EEG

Gen Hori*^, Takayuki Hamano[#+], Hidefumi Ohmura[$+], Ryu Nakagawa[%+], Kiyoshi Furukawa[#], Kazuo Okanoya[&+]
***Faculty of Business Administration, Asia University**
^Brain Science Institute, RIKEN
[#]Department of Intermedia Art, Tokyo University of the Arts
[$]Department of Information Sciences, Tokyo Denki University
[%]Graduate School of Design and Architecture, Nagoya City University
[&]Graduate School of Arts and Science, The University of Tokyo
[+]JST-ERATO Okanoya Emotional Information Project

Our research aims at the retrieval of imagined musical chord from EEG recordings. We conducted an experiment using a 32 channel EEG device and 22 subjects (10 males and 12 females) who listened to and imagined five predetermined musical chord progressions. We found that, for all the subjects, the power of the alpha band (around 10Hz) significantly decreased at the last chord but one (V of the V-I motion or cadence). In other words, we observed alpha band desynchronization of EEG at the last chord but one at cadence. We made 160-dimesional feature vectors from EEG recording segments and classified them according to the chord (V or I) the subject imaged using SLR (Sparse Logistic Regression) and SVM (Support Vector Machine) to see that SLR outperforms SVM in the classification of imagined musical chord from EEG recordings.

Using representational similarity analysis to study perception of tonal categories

Blair Kaneshiro*, Tom Collins^, Anthony M. Norcia[+], Jonathan Berger*
***CCRMA, Stanford University**
^Music, Technology and Innovation Research Centre, De Montfort University
[+]Department of Psychology, Stanford University

Listeners form categories implicitly or explicitly as a result of exposure to music (Deutsch 1980; Tillmann 2000). Here we use Representational Similarity Analysis (RSA) to investigate perception and representation of tonal categories. RSA facilitates the study of stimulus structure across response modalities (e.g., neurophysiological/behavioral responses, computational/theoretical models) by analyzing pairwise dissimilarities between responses, as opposed to the responses themselves (Kriegeskorte 2008). In a previous study, we classified single EEG trials of responses to culminating events in cadential formulae, including the expected tonic, anomalous chords (dominant, flatted supertonic) and silence (Kaneshiro 2012). We now perform RSA on the resulting confusion matrices using a recently developed technique (Kaneshiro 2015), and find that tonic (stable) and silent (highly unexpected) events each form tight, distinct category clusters across keys, while the remaining (unstable) events form a single category cluster. We supplement these results using computational models at the intersection of music psychology and MIR to attempt to simulate formation of the observed categories (Collins 2014). This RSA approach, which unites single-trial EEG responses and MIR models, elucidates categorical perception of harmonic classes. We propose that the approach will enable cortical investigations of perception of functional tonal relationships throughout entire harmonic progressions.

Multiple F0 Estimation by Gradient Frequency Neural Networks

Ji Chul Kim & Edward W. Large
Music Dynamics Laboratory, University of Connecticut

Estimation of multiple fundamental frequencies (F0s) from a mixture signal is an important aspect of computational auditory scene analysis and has a wide range of applications in music informatics research. Conventional methods for F0 estimation feature signal processing techniques, such as harmonic template matching and autocorrelation, for which biological evidence is largely lacking. Here we present a novel method for F0 estimation based on the nonlinear dynamical processes found in the auditory system. We use a canonical model for gradient frequency neural networks, a mathematically simple yet biological plausible model of tonotopically organized networks of neural oscillators. We show that spatiotemporal patterns of synchronous oscillations emerging in response to complex sounds can serve as a reliable F0 estimator. When driven by a mixture of multiple harmonic sounds, oscillators tuned to the harmonics of each sound form a segregated pattern of mode locked synchronization, which provides an estimate of F0 for each sound. We evaluate the performance of the model using mixtures of the recordings of musical instruments. Results show that our approach represents a biologically realistic, yet viable alternative to existing methods for multiple F0 estimation.

Adaptivity in oscillator-based pulse and metre perception

Andrew J. Lambert, Tillman Weyde, Newton Armstrong
City University London

Beat induction is the perceptual and cognitive process by which we listen to music and perceive a steady pulse. Computationally modelling beat induction is important for many MIR methods and is in general an open problem, especially when processing expressive timing, e.g. tempo changes or rubato. Large et al. (2010) have proposed a neuro-cognitive model, the Gradient Frequency Neural Network (GFNN), which can be model the perception of pulse and metre. GFNNs have been applied successfully to a range of 'difficult' music perception problems (see Angelis et. al., 2013; Velasco and Large, 2011). We have found that GFNNs perform poorly when dealing with tempo changes in the stimulus. We have thus developed the novel Adaptive Frequency Neural Network (AFNN) that extends the GFNN with Righetti et al.'s (2006) Hebbian learning rule. Two new adaptive behaviours (attraction and elasticity) increase entrainment, and increase model efficiency by allowing for a great reduction in the size of the network. We will extend our comparative study presented at this year’s ISMIR with a new set of results incorporating different forms of expressive timing and we will discuss the biological and cognitive plausibility of the AFNN.

The role of surprise in music

Scott A. Miles*, David S. Rosen^, Norberto Grzywacz**
***Interdisciplinary Program in Neuroscience, Georgetown University**
^Music & Entertainment Technology Lab, Drexel University
****School of Arts & Sciences, Neuroscience and Physics, Georgetown University**

Does something about the harmonic structure of music drive preference? Information theory might help us understand the relationship between the ordering of chords and how music is perceived. We analyzed chord transcriptions from the McGill Billboard Project corpus to determine whether surprise plays a role in the success of a song. In our analyses, "surprise" increases as the probability of a specific event occurring, given the prior context, decreases. We measured the average surprise along songs in the top and bottom quartiles of the corpus, according to peak Billboard chart position. We also measured the surprise during various sections of songs, including "verse", "chorus", and "bridge", in the top and bottom quartiles. We found that average surprise along the trajectory of top quartile songs is consistently higher. We also found that much of the variability in surprise between quartiles may be accounted for by differences in the verse sections – especially in the verse sections that immediately precede a chorus. These findings are consistent with two hypotheses: First, higher overall surprise leads to more interest, and therefore higher preference. Second, higher localized surprise immediately preceding a lower surprise section leads to a contrast between tension and release, and therefore higher preference.

Pattern discovery and music similarity with compositional hierarchical model

Matevž Pesek*, Aleš Leonardis, Matija Marolt***
***University of Ljubljana, Faculty of Computer and Information Science**
****University of Birmingham, School of Computer Science**

We present a compositional hierarchical model for symbolic music representations (SymCHM). The SymCHM is a deep architecture model with a transparent multilayer structure. The model can be used for pattern discovery, as well as music similarity. It can learn a set of representative repeated patterns of individual works or larger corpora in an unsupervised manner, relying on statistics of pattern occurrences. A learned model contains representations of patterns on different layers, from the less complex on lower layers to the longer and more complex on higher layers. Its transparent nature enables insight into the found patterns, while the inference process with hallucination and inhibition mechanisms enables the search for pattern variations. The SymCHM is general in the way it represents the learned patterns and can be employed for various tasks, including pattern finding, similarity and novelty estimation and composer identification. For evaluation, we focused on the Discovery of repeated patterns & sections task proposed by Mirex community. We evaluated the model in the Mirex 2015 evaluation campaign and on the JKU PDD dataset.

Blue notes in Rock: An exploratory study

Iris Yuping Ren, David Temperley, Zhiyao Duan
University of Rochester

"Blue notes" are notes that fall in between chromatic scale categories; they are especially common between 3 and b3. While several authors have asserted that blue notes occur in rock, this has never been empirically verified. In this paper we present an exploratory study of blue notes in rock, focusing on the Jackson 5's "ABC". This song was chosen because a recording of the isolated vocal is available and because it appears to use both 3 and b3 quite extensively. We analyzed the pitch contour of the isolated vocal using a pitch-tracking algorithm, and calculated the mean pitch of each mediant note (in the neighborhood of 3 or b3). We then examined each mediant note in context, seeking to explain why the singer chose 3, b3, or something in between. (Various factors guiding choices between 3 and b3 have been discussed in the rock literature.) We found several notes that seemed like plausible candidates for intentional blue notes; these were all cases in which the contextual factors seemed ambivalent between 3 and b3. While the current study is only exploratory, we believe it could provide a useful starting point for future research on blue notes.

Neural resonance and beat tracking in complex meters

Carson G Miller Rigoli*, Sarah Creel*, Petr Janata**
***University of California, San Diego**
****University of California, Davis**

Compared with industry algorithms for beat-tracking, neural resonance models offer psychologically plausible mechanisms, which may provide more accurate simulations of human musicality. Unfortunately, quantitative data of human beat tracking is often lacking and even cognitive models must rely on musicological interpretations of beat and meter. This is especially the case for complex meters. Here, human tapping data is collected to replicate findings indicating that Westerners systematically distort beat times from normative beats in complex meters such as 7/8. Beat time predictions and measures of distortion for 7/8 meters are also calculated using two neural resonance models of beat-tracking, a Gradient Frequency Neural Network (GrFNN) and the Beyond the Beat model (BTB). GrFNN excelled at finding subdivision-level periodicities but was unable to extract the target beats. GrFNN also did not produce distortion at the target beats in line with Western performance. BTB on the other hand was able to extract beats in line with normative musicological descriptions. However, BTB also failed to produce distortion as seen in real tapping data. This study serves as an example where traditional musicological understandings fail to capture real human perception and underscores the need for real performance data to create MIR applications with more human-like performance.

Validating a technique for post-hoc estimation of a listener's focus in music structure analysis

Jordan B. L. Smith*, Elaine Chew**
***National Institute of Advanced Industrial Science and Technology (AIST), Japan**
****Queen Mary University of London**

Research on music structure analysis may be hindered by reductive annotations that are limited to boundary locations and section labels. This research could be significantly advanced by knowledge of the reasoning behind these analyses: what musical patterns justified the decisions of the listener? What was the listener likely to have been paying attention to? An algorithm, based on quadratic programming, for estimating which musical features formed the basis of a listener's analysis was previously proposed. Here, we verify the approach by testing it on a set of artificial musical stimuli created to have controlled musical changes, which were validated by listeners who were instructed to focus on particular features. We find that the proposed algorithm is in fact able to predict, above chance, the feature to which the listeners were paying attention. We also test a few variants of the algorithm, and find that a correlation based approach works best. However, the high error rate (36%) suggests that the link between qualitative features like "timbre" and quantitative features like "MFCCs" can be surprisingly weak, and perhaps erroneously taken for granted.

Learning about music cognition by asking music information retrieval questions

Sebastian Stober
Machine Learning in Cognitive Science Lab, University of Potsdam, Potsdam, Germany

Since its publication at ISMIR 2015, the OpenMIIR dataset of EEG recordings taken during music perception and imagination has been analyzed in several experiments. In these experiments, typical questions from music information retrieval (MIR) were addressed such as: Can we recognize the individual music pieces from the EEG, track their tempo, identify the meter or the presence of a singing voice, or even reconstruct the original audio stimuli? Developing solutions for such problems has become an interesting challenge for the MIR community. At the same time, the results from this research have begun to also lead to new insights and novel questions in the field of music cognition. For instance, the 12-class stimulus identification problem can be addressed by training a very simple convolutional neural network with just a single convolutional filter aggregating the 64 EEG channels and a fully connected layer that learns a temporal pattern for each stimulus. Training such a model across subjects leads to an accuracy of over 40% for individual subjects while the temporal patterns reveal that the network is able to capture musically meaningful events such as down-beats. This talk will highlight findings from several such MIR experiments and discuss possible implications.

Identification of listener genre preference using the Echo Nest API features

Naresh Vempala*, Michael Barone, Matthew Woolhouse****
***SMART lab, Ryerson University**
****McMaster University**

This research compared the performance of four different machine learning (ML) models trained to correctly classify people’s genre preferences using the audio features of tracks within their download collections. The study’s data, taken from a database of music downloads made by users of Nokia mobile phones between 2007-14, consisted of 1200 listeners, each belonging to one of six X-head subgroups, where X was the most numerous genre within their collection (e.g. Metal-head, Pop-head). Each Xhead subgroup contained 200 representative listeners with 100-1000 downloads within their respective collections. Data pre-processing involved calculating an average audio feature vector containing ten features per listener, such as acousticness, danceability, and energy, obtained for each track via The Echo Nest API. Ten-fold cross-validation was used to train and test the four ML models. The goal for each model was to predict an X-head’s genre preference given his or her average feature vector. Of the four ML models - Support Vector Machines, Random Forests, k-Nearest Neighbours, and Linear Discriminant

Analysis - results showed that Random Forest and Linear Discriminant models outperformed others. We discuss implications of these results and the relevance of the audio features within our study with respect to current ML literature and research.

A probabilistic model of meter perception

Bastiaan van der Weij*, Marcus Pearce, and Henkjan Honing***
***Institute for Logic, Language and Computation (ILLC), University of Amsterdam**
****School of Electronic Engineering and Computer Science, Queen Mary University of London**

Research shows that enculturation affects meter perception. Predictive coding posits that brains pursue a minimization of prediction error through predictive models tailored by previous experience. We hypothesize that inference of a metrical structure for interpreting rhythms forms part of a prediction-error reduction strategy. Since the underlying predictive model is tailored by previous experience, this may explain why listeners with different cultural histories of exposure to rhythm show differences in their metrical interpretation of rhythms. We propose a probabilistic model of meter perception employing an empirical Bayes scheme to learn from an annotated dataset of rhythms, and infer a distribution over metrical interpretations for unseen rhythms. The model is formulated as an incremental model that predicts the timing of events in a rhythm based on events observed so far and inferred metrical categories. We operationalize prediction error as information content to measure how well the model predicts previously unseen rhythms. We measure the average likelihood the model assigns to annotated metrical interpretations to assess how well the model's output matches annotator's intuitions. In a proof-of-concept simulation of enculturation, we train the model on datasets containing stylistically different sets of rhythms and investigate how training data impacts the model's interpretation of rhythms.