August 8, 2019

Brooklyn College, Tow Center for the Performing Arts
Brooklyn, NY

Schedule:

09:00 am-09:35 am	Registration
09:35 am-09.45 am	Welcome Address (Naresh Vempala/Frank Russo/Johanna Devaney)
09:45 am-10:30 am	Opening Keynote – **Stephen McAdams:** *A framework for modeling perceptual effects of auditory scene analysis in orchestral scores*
10:30 am-10:50 am	Break
10:50 am-11:50 am	**David Baker:** *Modeling the limits of musical memory* **Douglas Scott:** *Violations of Zipf's law as a tool for modelling and visualizing musical structure* **Ana Clemente, Manel Vila-Vidal, Marcus Pearce, Marcos Nadal:** *A set of 200 musical stimuli varying in balance, contour, symmetry, and complexity: behavioral and computational assessments* **Aurélien Antoine, Philippe Depalle, Philippe Macnab-Séguin, Stephen McAdams:** *Identifying orchestral blend effects from symbolic score data*
11:50 pm-12:25 pm	**Speed poster talks** (see page 2)
12:25 pm-01:25 pm	Lunch/Poster session
01:25 am-02:10 pm	Keynote – **Claire Arthur:** *Bringing more information to music informatics: Combining tools, data, and best practices from cognition and MIR*
02:10 pm-03:10 pm	**Michael Schutz:** *New software for exploring temporal structure* **Ji Chul Kim:** *Periodicity detection in syncopated rhythms: nonlinear oscillator vs. standard methods* **Micael Antunes, Danilo Rossetti, Jônatas Manzolli:** *Perception of sound mass and an emergent harmony in György Ligeti's Continuum* **Ilana Harris & Mats Küssner:** *Predicting emotion through body language: improved performance in Generative Adversarial Networks trained on unlabeled video of musical group interaction*
03:10 pm-03:30 pm	Break
03:30 pm-04:00 pm	**Simin Yang, Mathieu Barthet, Elaine Chew:** *Listener-informed features for time-varying emotion perception in live music performance* **Pranav Sankhe:** *An information theoretical approach towards the reconstruction of tempo from EEG responses*
04:05 pm-04:50 pm	Closing Keynote – **John Ashley Burgoyne:** *Everyday features for everyday listening*
04:50 pm-05:20 pm	Announcement of Student Prizes/Closing Remarks

Speed poster talks:

1. **Dylan Scott Mandel & Nancy Eng**
   The added value of musical training on linguistic syntax processing
2. **Noah Kahrs**
   Higher chords are more closely spaced in semitones but not in Hertz
3. **Vincent Lostanlen, Joakim Andén, and Mathieu Lagrange**
   Learning auditory similarities between instrumental playing techniques
4. **Caitlyn Trevor**
   The cultural and ethological significance of tempo in suspenseful music
5. **Saebyul Park, Halla Kim, Juyong Park, Jeounghoon Kim, Juhan Nam**
   SMPD: Symbolic Music Plagiarism Dataset
6. **Matthew Moreno & Earl Woodruff**
   Integration of emotional and psychophysiological tools to examine a listener’s response to musical stimuli while completing a reading comprehension task

Partners

SCIENCE of music AUDITORY research & technology

SMPC
Society for Music
Perception and Cognition

Brooklyn College

Sponsors

LUCID

NSERC
CRSNG

Keynote Addresses:

Stephen McAdams, McGill University

A framework for modeling perceptual effects of auditory scene analysis in orchestral scores

Music uses the qualities of different instruments to create specific perceptual, expressive, and emotional effects that composers sculpt over time. Timbre perception is multifaceted and contributes in many ways to the perceptual organization of musical structures. Orchestration taken its broadest sense is the conception, selection, combination, and juxtaposition of sound qualities to achieve a specific musical aim. The aim of this work is to develop a theoretical ground for orchestration practice starting with the structuring role that timbre can play in music, using scene analysis as a starting point. Many facets of musical structuring are achieved by auditory scene analysis, the perceptual grouping processes that: 1) fuse different acoustic components into events (e.g., instrumental blend), 2) integrate events into one or more auditory streams or other sequential groupings (e.g., surface textures or orchestral layers), 3) segment groups of events into motifs, phrases, and sections (e.g., antiphonal contrasts, section boundaries), and 4) form larger-scale units encompassing changes in orchestration that are extended over time (e.g., orchestral gestures). The roles that timbre plays in the manifestation of these perceptual principles in orchestration practice will be considered as a point of departure for modeling them based on both symbolic score data (MusicXML) and signal properties.

Claire Arthur, Georgia Institute of Technology

Bringing more information to music informatics: Combining tools, data, and best practices from cognition and MIR

Despite being the two main branches of music-scientific research, music cognition and music information retrieval (MIR) largely remain independent fields. Yet, each field has the potential to strengthen the other. In this talk I will contrast approaches to experimental design, data gathering, analysis, and evaluation across both music cognition and music information retrieval. By highlighting similarities and differences, and exploring the optimal practices of each domain, I will offer suggestions for concrete ways in which each field can augment efficiency, reproducibility, and motivation by borrowing ideas, data, and techniques from the other. Drawing on examples from recent projects in melodic expectancy, key-finding, and automatic harmonic analysis, I aim to highlight ways that taking advantage of tools and practices in each field can help “bridge the divide” and unify research across domains.

John Ashley Burgoyne, University of Amsterdam

Everyday features for everyday listening

You are sitting on a commuter train. How many passengers are wearing headphones? What are they listening to? What else are they doing? Most importantly, amid the cornucopia of distractions, what exactly are they hearing? Much research in music cognition pits ‘musicians’, variously defined, against non-musicians. Recently, especially since the appearance of reliable measurement instruments for musicality in the general population (e.g., Müllensiefen et al., 2014), there has been growing interest in the space in between. Moreover, the ubiquity of smartphones has greatly enhanced the ability of techniques like gamification or Sloboda’s experience sampling to reach this general population outside of a psychology lab. MIR can provide the last ingredients to understand what is happening between our commuters’ earbuds: everyday features for studying everyday listening. Since Aucouturier and Bigand’s 2012 manifesto on the poor interpretability of traditional signal-processing measures, clever dimensionality reduction paired with feature sets like those from the FANTASTIC (Müllensiefen and Frieler, 2006) or CATCHY (Van Balen et al., 2015) toolboxes have sought a middle ground.

This talk will present several uses of everyday features from the CATCHY toolbox for studying everyday listening, most notably a discussion of the Hooked on Music series of experiments (Burgoyne et al., 2013) and a recent user study of thumbnailing at a national music service. In conclusion, it will outline some areas where MIR expertise can go further than just recommendation to learn about and engage with listeners during their daily musical activities.

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

Identifying orchestral blend effects from symbolic score data

Aurélien Antoine, Philippe Depalle, Philippe Macnab-Séguin, Stephen McAdams
Schulich School of Music, McGill University

We report on an ongoing research project that aims to computationally model perceptual effects of orchestration using symbolic, audio, and perceptual information. Here, we focus on orchestral blend, which happens when sounds coming from two or more instruments are perceived as a single sonic stream. In this first phase, we developed models to identify orchestral blend effects from symbolic information taken from scores. Several studies have suggested that different musical properties contribute to create such effects. The blend estimations of our models are based on calculations related to onset synchronicity, pitch harmonicity, and parallelism in pitch and dynamics, using symbolic information contained in MusicXML files taken from the OrchPlayMusic Library (orchplaymusic.com). In order to assess the performances of the models, we applied the models to different orchestral pieces and compared the outputs with human experts' ratings in the Orchestration Analysis and Research Database (orchard.actor-project.org). Using different weights for the three parameters, the models obtained an average accuracy score of 75%. These preliminary results support the initial developments and suggest that estimations based on symbolic information would account for a significant part in modeling orchestral blends. Future work aims to include audio analyses to take into account timbral properties as well.

Perception of sound mass and an emergent harmony in György Ligeti's Continuum

Micael Antunes, Danilo Rossetti, Jônatas Manzolli
Interdisciplinary Nucleus for Sound Communication, Institute of Arts, University of Campinas

The goal is to present a methodology of analysis to describe the sound masses from a psychoacoustic standpoint, analyzing the work Continnum (1968) of György Ligeti. Ligeti's instrumental music in such period explores the fusion of sound by experimenting different sound sequences superimposed, creating a complex polyphony, known as timbre of movement (Ligeti, 2010: 185). We apply a model of psychoacoustic audio descriptors to the audio recording, intending to reveal emergent features of the sound texture which are beyond musical notation. The representations based on the descriptor's data extracted (Figure 1) are compared with a listening and traditional score analyzes. Bark coefficients (Bullock, 2007) reveal that all critical bandwidths are stimulated along the piece, varying in different moments. Chroma (Mauch, Dixon, 2010), gives similar information, but in relation to pitches. Spectral flux (Peeters, 2004) describes the intensity variation between two successive frames. The volume descriptor (Malt, Jourdan, 2009) defines the frequency and intensity space that the texture occupies. By observing the graphics, the excerpt between 1'48'' and 2'30'' is the climax of the piece in terms of intensity, flux and densification of the texture. The present descriptors allow to explore nuances and specificities of the sound masses in Continuum. Parameters as density, energy, frequency range and overlap of sound masses will be examined in more detail in the full paper.

Modeling the limits of musical memory

David Baker
Louisiana State University

We present both evidence from a newly encoded corpus of over 783 sight singing melodies and a currently ongoing experiment (N = 11, Stopping = 75) in order to answer this question. Using a within-subjects design and musical series recall task using stimuli taken from our corpus with trained musicians, this paper will attempt to explain the results from the behavioral experiment using various models that have been put forward by both the computational musicology and cognitive psychology literature. Models we include range from number of items/notes (Miller, 1956; Tallarico, 1974; Long, 1977; Pembrook, 1983; Li, Cowan, Saults, 2012) , to an n-grams frequency distribution (Huron, 2006), to information content measures derived from multiple viewpoint models of music perception (Pearce, 2005; Witten and Conklin, 1995), sensory models of musical perception (Leman, 2000; Milne, Laney and Sharp, 2015), to static symbolic features (Müllensifen, 2009). We fit multiple models to our data predicting both accuracy and reaction time then discuss our results in terms of plausible cognitive explanations and discuss how insights from cognitive psychology can lead to more meaningful and robust models of musical memory.

A set of 200 musical stimuli varying in balance, contour, symmetry, and complexity: behavioral and computational assessments

Ana Clemente[1*], Manel Vila-Vidal[2], Marcus T. Pearce[3,4], & Marcos Nadal[1]

[1]Human Evolution and Cognition Research Group, University of the Balearic Islands, Institute for Cross-Disciplinary Physics and Complex Systems, Associated Unit to CSIC, Palma, Spain
[2]Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain
[3]School of Electronic Engineering & Computer Science, Queen Mary University of London, UK
[4]Centre for Music in the Brain, Department of Clinical Medicine, Aarhus University, Denmark

We introduce a novel set of 200 Western tonal musical stimuli (MUST) for research on perception and valuation of music. The set consists of four subsets of 50 4-s motifs varying in balance, contour, symmetry, or complexity. They were designed to be both musically appealing and experimentally controlled. We assessed them behaviorally and computationally. The behavioral assessment aimed to determine whether musically untrained participants could identify variations in each attribute. The results showed high inter-rater reliability and that ratings mirrored the design features well. Participant's scores also served to create an abridged set. The computational assessment required developing a specific battery of computational measures that describe each stimulus in terms of its structural parameters. The distilled non-redundant composite measures proved to be excellent predictors of participants' ratings, and the complexity composite measure resulted better or as good as existing models of musical complexity. The MUST set and computational measures as a package of functions for MATLAB are suitable for use in studies that require presenting short motifs varying in balance, contour, symmetry, or complexity. They are valuable resources for research in music psychology, empirical aesthetics, music information retrieval, musicology, etc., freely available through the Open Science Framework and GitHub.

Predicting emotion through body language: improved performance in Generative Adversarial Networks trained on unlabeled video of musical group interaction

Ilana Harris & Mats Küssner
Humboldt-Universität zu Berlin, Berlin, Germany

Musical group interaction, where two or more individuals play music together, is universal to human experience, and is a crucial mechanism of communication, social bonding, and cultural transmission (Tarr et al., 2014). Prior research has revealed increased levels of empathy after implementing long-term musical group interaction paradigms in primary school children (Rabinowitch et al., 2012). In this study, we examine musical group interaction in global live electronic dance music as fostering increased empathy through communication of emotion through body language. We train Generative Adversarial Networks (GAN) on unlabeled video documenting musical group interaction in cross-cultural instances of live electronic dance music. Preliminary results show improved performance in the model's ability to categorize emotional valence based on body language after training with these clips in contrast to control videos of group interaction of spectators at live sports matches. Implications include a) encouraging big data research within music psychology to facilitate research of universals related to high-level social processes b) increased understanding of the effects that growing corpuses of big music data have on cross-cultural emotional experience and c) aiding in the development of public policy in contributing to policymakers' decisions regarding copyright law of data mining digital multimedia.

Higher chords are more closely spaced in semitones but not in Hertz

Noah Kahrs
Eastman School of Music, University of Rochester

Computational models of tonality tend to assume relative pitch; key profiles, for instance, measure scale-degrees against a mobile tonic (Huron, 2006; Temperley & Marvin, 2008). However, absolute pitch effects disrupt tonal perception via both key characteristics (Quinn & White, 2017) and absolute boundaries in register (Biasutti, 1997). Consequently, chords might not have uniform properties under transposition. Huron (2001, 2016) has noted that chords with higher bass notes span fewer semitones, but he does not provide statistical detail. In this study, I present a corpus analysis of chord spacings in 60 Haydn and Mozart string quartet movements, specifically of 2,502 triads and tetrachords therein. Chords with higher bass notes are indeed more closely spaced not just in terms of total span but also in mean and median distance between notes, as long as the distance is measured in semitones. However, these correlations disappear when pitches are instead encoded exponentially in Hertz, as is done in spectrum-oriented studies (Peeters et al., 2011). Chords thus have a consistent distribution of frequency spans, significantly more than in a control corpus randomly generated from the instruments' pitch distributions. Even without reference to instrument spectra (Huron & Sellmer, 1992), different pitch scalings yield different conclusions.

Periodicity detection in syncopated rhythms: nonlinear oscillator vs. standard methods

Ji Chul Kim
Oscilloscape, LLC; University of Connecticut

Periodicity detection is an essential step in the retrieval of rhythmic information such as tempo estimation and beat tracking. Standard signal processing methods, including autocorrelation, comb filtering and the Fourier transform, have been used in the literature to estimate periodicities in music signals. Here we study periodicity detection by nonlinear oscillators and compare it with the standard methods. Our focus is the periodic pulse or beat in syncopated rhythms, and we examined whether the output of each method matches the findings in human experiments. Mathematical analyses and discrete-time computations with audio signals showed that oscillators with a nonlinear input function exhibit multistability and hysteresis found in the human data. Individual oscillators phase-locked to one of the potential beat positions perceived by human listeners, and the beat selection by a population of oscillators increased multiplicity as rhythmic complexity increased, matching the trend in the human data. Other methods either did not show multistability (the Fourier transform), represented multiple phases in single units (comb filters), or did not include phase information (autocorrelation). This shows that nonlinear oscillators behave more like individual human listeners and thus may have advantages over other methods in detecting rhythmic periodicities, especially for highly syncopated complex rhythms.

Learning auditory similarities between instrumental playing techniques

Vincent Lostanlen[1], Joakim Andén[2], Mathieu Lagrange[3]

[1]New York University
[2]Flatiron Institute
[3]CNRS / University of Nantes

The modeling of timbre is typically based on measuring human responses to predefined stimuli, such as musical instruments and speech. However, the particular stimuli, such as musical notes, are often too simple to explore the complex structure of timbre. Other stimuli, such as speech sounds, provide a richer timbre, but also carry higher-level meaning, making unbiased responses harder to obtain. Between these extremes, we consider the use of musical instruments coupled with various instrumental playing techniques (IPTs), offering a broader range of timbre than ordinary notes (i.e., played ordinario), while largely avoiding semantic connotations. We show how human subjects organize these instrument-IPT pairs in free sorting tasks into clusters that suggest a more general taxonomy than that provided by instrument or IPT alone. In addition, we propose a computational model for reproducing this timbral organization based on joint scattering transforms connected to an unsupervised linear layer. This model shows good agreement with the timbral similarity judgments of the subjects, providing a promising approach for further exploration of this perceptual dimension.

The added value of musical training on linguistic syntax processing

Dylan Scott Mandel, Nancy Eng
CUNY, Hunter College School of Health Professions

Long-term musical training has been shown to improve a variety of linguistic functions. Certain aspects of music are akin to patterns and rules that have been compared to grammatical rules in language. Previous research supports this notion by demonstrating that language and music draw on shared resources to process incoming information. However, whether the amount of experience and/or expertise with music play any role in language processing has yet to be explored. The present study explores this notion of shared resources for the processing of language and music with two groups of individuals that are defined as musicians and nonmusicians. Participants performed self-paced reading of unexpected linguistic structures (e.g., garden path sentences) paired with unexpected harmonic structures (e.g., out-of-key chords). Results demonstrate musicians showed reduced garden path effects (i.e., shorter reading times), while nonmusicians showed increased interpretation difficulties with accompanying slow resolution. These results strengthen a limited line of research which points to the processing of linguistic and musical syntax differing with musical ability, and provide further evidence for the shared processing of their mechanisms.

Integration of emotional and psychophysiological tools to examine a listener's response to musical stimuli while completing a reading comprehension task

Matthew Moreno & Earl Woodruff
Ontario Institute for Studies in Education, University of Toronto

Literature as explored the perceptual effects of musical tempo (Linek, Marte, & Albert, 2011; Thompson, Schellenberg & Husain, 2016) as they pertain to the cognitive controls and emotions that we exhibit, but often examine these through a singular dimension. A multimodal examination of the perceptual effects of tempo are needed to understand the mechanisms that act upon human cognitive systems. To exploring this construct, a few questions were asked: 1) Is there a relationship between emotional expressions and performance on reading comprehension tasks, 2) Is there a relationship between emotional expressions and skin conductance levels across varying tempi, and 3) Do the slopes of skin conductance vary over time? In this study, 1st year undergraduate students (N=50) at a research-university were recruited to complete a reading comprehension task. Multiple reading samples were presented with 3 musical conditions played in the background: 1) no music (control), 2) slow music, and 3) fast music. Participants were monitored using facio-muscular emotion recognition software (iMotions Emotient) as well as electrodermal analysis (Biopac MP150). Through a developing a multilevel model (MLM), this study describes the interrelationship between emotional and psychophysiological responses and cognitive processes that predict the relationship between these underlying mechanisms. These results inform educators and music scientists about the multimodal antecedents of music on the cognitive process.

SMPD: Symbolic Music Plagiarism Dataset

Saebyul Park, Halla Kim, Juyong Park, Jeounghoon Kim, Juhan Nam
Korea Advanced Institute of Science and Technology

The purpose of this work is to build a symbolic melody dataset to facilitate musical analysis and interpretation of melodic similarity in the plagiarism context. The procedure of the study is as follows: 1) Case investigation and selection: plagiarism cases related to the melody are selected through investigation of copyright infringement cases.2) Melody transcription: the melody of selected songs is transcribed into MIDI based on the audio and score data. 3) Psychological experiment: an implicit memory task regarding similarity and memory is performed to see if the two melodies of the infringement case confuse the participants. Human annotated emotions (arousal and valence) and preferences are also collected during the experiment. As a result, a symbolic plagiarism data set (SMPD) containing information about the similarities, memories, emotions, and preferences of 132 songs in the copyright infringement suits from 1948 to 2017 is released.

An information theoretical approach towards the reconstruction of tempo from EEG responses

Pranav Sankhe
Indian Institute of Technology, Bombay

Here we propose an information theoretical model to analyze the reconstruction of the tempo of music stimuli from EEG responses. We interpret the entire transmission chain comprising of the stimulus generation, brain processing by the human subject, and the EEG response measurements as a nonlinear, time-varying communication channel with memory. We use mutual information (MI) to access the amount of information transfer from the music stimulus to the EEG response. We model the recorded EEG measurements as a multidimensional Gaussian mixture model (GMM). The input to our channel is the tempo value(X) which is modeled as a uniform random variable, and the output is the recorded EEG potential(Y) which is modeled as a GMM. The MI between the output EEG data and the tempo value tells us the maximum rate of change of tempo which we can afford in a music stimulus for perfect reconstruction of the tempo sequence.

The calculation of MI is $I(X; Y) = h(Y) - h(Y|X)$ .... (1)

where, $h(Z)$ is the entropy of the random variable Z.

We use Stanford's Naturalistic Music EEG Dataset - Tempo (NMED-T) to perform our computations of MI. To ensure the tractability of the MI computations and effective mapping of activity across different regions of the brain, we group the 128 electrodes of the EEG-system into 9 specific regions of interest (ROI). The obtained MI value averaged over all the stimuli was 20.83. This implies that for total reconstruction of the tempo sequence from EEG responses, the tempo value should not change faster than 20.83. This preliminary research establishes that using information theory, we can comment over the nature of input stimulus and establish bounds within which the reconstruction of the music stimulus is possible.

New software for exploring temporal structure

Michael Schutz
MAPLE Lab, McMaster University

Explaining the combined spectral and temporal complexity of natural musical sounds can be challenging when teaching auditory perception classes. Even single tones from musical instruments produce sounds that can only be fully conveyed using some type of 3D display showing continual changes in (1) amplitude (2) spectrum continuously over (3) time. It can be difficult to convey in textbooks figures, obfuscating the importance of dynamic temporal changes in acoustic structure and auditory perception. This pedagogical challenge mirrors a larger issue in auditory perception research, where the role of temporal changes in sounds often goes overlooked. This talk will review a software tool offering an intuitive framework for understanding important concepts related to timbre perception useful for students with a background in either music or the sciences. It allows students to explore and manipulate complex, time-varying sounds. The software uses an intuitive Graphical User Interface (GUI) to covey the role of temporal changes in amplitude. It can also be used to generate stimuli for perceptual experiments, facilitating better use of complex, time-varying sounds in the exploration of auditory perception. A beta version of this software is freely available to our colleagues and other interested parties at https://maplelab.net/pedagogy/

Violations of Zipf's law as a tool for modelling and visualizing musical structure

Douglas Scott
University of the Free State, South Africa

A minimum requirement for forming a conception of the world is to form a conception of boundaries that delimit some parts of it from others. It may be argued that the act of cognition is the act arranging perception into discrete partitions which, to the extent that they are stable under various types of manipulation, become the objects of experience. Using J.S Bach's Duetto BWV 802 as a concrete example, I describe a method for using Zipf's law as a systematic tool for investigating such stable boundaries and outline how this may be extended to other descriptive dimensions through a process called "semantic injection". While the musical object conceived of in this way exists in an infinite dimensional space which cannot be directly visualized, the result is surprisingly robust to even aggressive manipulation. This shows that even listeners with vastly different prejudices may nevertheless experience broadly the same object when listening to music, and that this object can be finitely, if imperfectly, described. The result can be useful for the purposes of music analysis, visualization and therapeutic applications, among others.

The cultural and ethological significance of tempo in suspenseful music

Caitlyn Trevor[1,2]

[1]**Ohio State University**
[2]**University of Zürich**

Horror films often have a musical theme for a stalker character, such as for the killer, Michael, in Halloween (1978). These "stalker themes" usually have a consistent, low beat that mirrors the slow, confident approach of the killer. Both the killer's gait and the tempo of their stalker theme convey a sense of inevitability and doom. What tempi do composers use to convey doom? How do these tempi compare to other natural and cultural tempi that could be associated with doom (e.g. average walking speeds, the tempi of the classical suspenseful topic ombra)? To investigate, 50 excerpts of 10 seconds in length will be taken from film and tv soundtracks that accompany scenes where characters are being stalked by a threat. Using the Python package Librosa, tempo and beat data will be taken from each excerpt. Next, the data will be analyzed in comparison with related tempi such as average walking & running speeds and tempi culturally associated with the theme of inevitability and doom. The main hypothesis is that suspenseful music contains tempi reflective of cultural and ecological signals of suspense and dread. This work contributes to research on ethological signals and musical topics in music and emotion research.

Listener-informed features for time-varying emotion perception in live music performance

Simin Yang, Mathieu Barthet, Elaine Chew
Queen Mary University of London

The purpose of this study was to gain a deeper understanding of the factors that influence time-varying emotion perception in music performance. A two-stage study was launched to investigate the reasons listeners produce different emotion annotations in a complex classical music piece. In an initial experiment, we collected time-varying emotion ratings (valence and arousal) from listeners of a live performance of a classical trio; In a follow-up experiment, we interrogated the reasons behind listeners' time-stamped emotion ratings, through the re-evaluation of seven pre-selected music segments of various agreement levels informed from the initial study. 15 and 21 participants with varying degrees of music training were involved in the initial and follow-up experiments, respectively. Thematic analysis of the time-stamped explanations revealed themes pertaining primarily to musical features of loudness, tempo, and pitch contour as the main factors influencing emotion perception. The analysis also uncovered features such as instrument interaction, repetition, and embellishments which are less mentioned in music informatics researches. With recent advances in MIR subtasks such as automatic instrument recognition, playing techniques detection and source separation, we propose the potential incorporation of these listener-informed features into MIR-based audio content analysis.