

GLAM 4: When Music Becomes Data: Modeling Structure, Culture, and Performance — Program

A workshop presented by Research Axis 2: Music Information Research

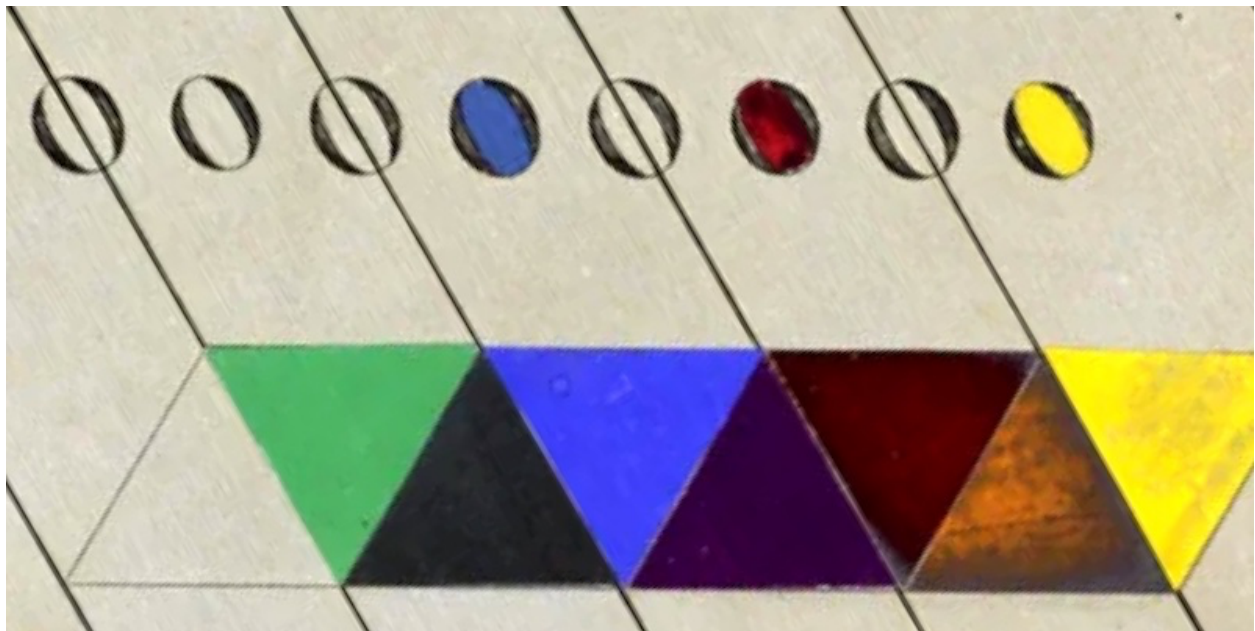
October 6, 2026, 9:30–12:00

Room A-832 (8th floor), Elizabeth Wirth Music Building, 527 Sherbrooke O

Description

Computational models of music can formalize assumptions about what music is, how it is structured, and what counts as musical knowledge when translated into computational form. This workshop invites contributions that examine how musical knowledge is represented, transformed, and interpreted across symbolic, audio, cultural, and performance contexts, foregrounding modeling as a potentially interpretive act rather than a neutral technical process.

The workshop follows Anja Volk's CIRMMT Distinguished Lecture, *Computational modeling of musical structures*, on October 5, 2026. Volk joins as a discussant, bringing her perspective on computational modeling of musical structures to bear on the contributions presented.



Hay, D. R. 1838. *The Laws of Harmonious Colouring: Adapted to Interior Decorations, Manufactures, and Other Useful Purposes*. London: W. S. Orr; Edinburgh: W. & R. Chambers. Plate, "Example I," 22. Getty Research Institute copy, digitized by Internet Archive. <https://archive.org/details/lawsofharmonious00hayd>.

Schedule

9:00 – 9:30 Coffee / Café

9:30 – 9:35 Welcome words / Mot de bienvenue

9:35 – 10:40 Individual talks / Présentations individuelles

10 minutes each, followed by a short question period

Zachary Sussman — *A Formal Grammar for Modeling the Experience of Arrival and Departure in Polyphonic Musical Events*

Paul Robert — *Tracking Sensory Predictions in Neural Responses to Naturalistic Music with Deep Learning Models*

Farzad Milani — *Encoding a Forgotten System: Hedāyat's Abjad-Metric Notation as a Model of Seventeen-Tone Persian Music*

Kyrie Bouressa — *Automatic Detection of Musical Scribes in Saint Gall Codices 390/1, 389, 340*

Yifan Huang & Eto Sun — *Encoding Guqin Jianzipu: A Gesture-Based MEI Module for a Logographic Notation*

10:40 – 11:00 Coffee break / Pause café

Demo Orfeo by Élie Simard

11:00 – 11:50 Individual talks / Présentations individuelles

Zih-Syuan Lin — *What Counts as a Correct Score? Musical Assumptions in Transcription Evaluation*

Hanwen Zhang — *A Context-Aware Piano Performance Encoding Framework for Sustain Pedaling*

Yashar Talebirad — *What Makes Sound Layers Clash? Human Judgments, Signal Features, and Audio Models*

Wenye Ma — *What Matters for Music-Centered Recognition in Audio-Language Models?*

11:50 Closing discussion / Discussion de clôture

Presentation abstracts

Zachary Sussman — *A Formal Grammar for Modeling the Experience of Arrival and Departure in Polyphonic Musical Events*

In my presentation, I will begin by explaining how current formal models of musical structure describe compositional content in terms of objective materials such as rhythm, meter, tempo, pitch, harmony, and sonic material. However, what these models often do not capture is how these materials are experienced by the listener. I will introduce my approach to exploring how formal models can instead describe aspects of the listener's experience itself.

I will present and define an abstraction which I refer to as the experience of arrival and departure. I will then introduce and demonstrate the syntax of a programming language I have developed for describing this experience. I will show how programs written in the language correspond to experiences of arrival and departure associated with defined polyphonic musical events, using several examples. Additionally, I will demonstrate the animation and sonification capabilities of my online IDE, which generates visual and auditory materials intended to elicit the modeled experience of arrival and departure.

Finally, I will present the next steps for this work in terms of how I plan on expanding the language to account for more aspects of musical experience that are not currently captured by formal descriptions of musical structure.

Paul Robert — *Tracking Sensory Predictions in Neural Responses to Naturalistic Music with Deep Learning Models*

In cognitive neuroscience, computational models of music are used to simulate how listeners perceive musical structures and to investigate the neural responses they evoke. More specifically, a common approach involves using models that predict upcoming pitches and rhythms in a musical sequence, and neural error signals reflect the unexpectedness of musical events to the listener. In recent years, Markov chain and deep learning models have yielded fruitful results at the neurophysiological level, but have only been applied to melodies or isochronous chord series. To progress towards more ecological paradigms, we aimed to extend this approach to naturalistic, polyphonic music. We retrieved an existing generative music model (Performance RNN) operating on MIDI tokens and trained a new model using a piano-roll representation (PolyRNN). By analysing neurophysiological activity acquired non-invasively (MEG) and within the human brain (intracranial EEG) while participants listened to expressive piano recordings, we show that piano-roll-based modeling better captures neural error signals in polyphonic contexts.

Farzad Milani — *Encoding a Forgotten System: Hedāyat's Abjad-Metric Notation as a Model of Seventeen-Tone Persian Music*

In the early twentieth century, Persian musicologist Mehdi-Qoli Hedāyat devised a hybrid Abjad-Metric notation system, building on a seventeen-tone Abjad tone-letter tradition dating back to thirteenth-century treatises. In this earlier system, rhythmic and metric values had never been notated with precision; Hedāyat's central innovation was to fuse the Abjad tone-letters with Western metrical time values to resolve this gap. Designed for the setār and free of staff, clefs, and accidentals, his system emerged at roughly the same historical moment as Ali-Naqi Vaziri's competing notation, which instead adopted fully European staff notation built on twenty-four-tone equal temperament, adding two accidental signs to represent quarter-tones. Where Vaziri's system prioritized legibility for musicians trained in European notation, Hedāyat's hybrid approach more faithfully preserved the microtonal nature of Persian music, at the cost of being largely illegible to musicians unfamiliar with the Abjad tradition. This presentation treats Hedāyat's notation not as a historical curiosity but as a case study in how notational choices formalize assumptions about pitch, structure, and musical knowledge. By reconstructing and transcribing Hedāyat's only surviving song-text example, the analysis surfaces the interpretive decisions embedded in translating a modal, orally-transmitted tradition into a fixed symbolic system, and asks what is gained and lost when non-tempered, culturally specific tuning systems are mapped onto standardized notational or computational frameworks.

Kyrie Bouressa — *Automatic Detection of Musical Scribes in Saint Gall Codices 390/1, 389, 340*

Scribal identification in medieval manuscripts enables scholars to trace networks of production, circulation, and institutional exchange. Machine learning approaches to scribal identification in medieval text have advanced significantly, while musical notation remains underexplored.

This paper presents a methodology for detecting scribal changes in musical manuscripts using learned feature representation and nearest-neighbor comparisons. In a St. Gall case study, features derived from a self-supervised model form stable clusters that correspond to scribal transitions.

Against SG 390, the approach recovers two profiles consistent with a mid-volume change of musical scribe. A test on SG 389 supports a common musical hand with one of these profiles, while SG 340—used as a true negative—does not. This research contributes a framework for digital musical paleography that complements traditional codicological methods and demonstrates how digital tools can assist in detecting patterns in musical scripts, production networks, and the circulation of musical knowledge across institutions.

Yifan Huang & Eto Sun — *Encoding Guqin Jianzipu: A Gesture-Based MEI Module for a Logographic Notation*

We propose MEI.jianzipu, the first Music Encoding Initiative (MEI) module for a non-pitch-centric, gesture-based East Asian notation system: jianzipu. Jianzipu (减字谱, “reduced-character notation”) is a logographic system that encodes fingering and playing gestures rather than abstracting pitch. With a history of approximately 1,300 years and continued use in contemporary practice, it remains the primary notation for the guqin, the Chinese seven-string zither and one of China’s foremost solo instrumental traditions.

Jianzipu differs fundamentally from Western Common Music Notation (CMN), which treats the note as an atomic unit defined by pitch and duration. In jianzipu, by contrast, the primary unit is the physical action—touch type, movement direction, finger identity, and positional information—from which pitch is a derivative outcome. This allows the notation to encode the micro-gestural and timbral nuances central to guqin performance. Likewise, jianzipu diverges from Western string tablature, which relies on grid-based string-fret coordinates as a pitch surrogate and seldom captures the gestural morphology of sound production. Despite its musical and historical significance, jianzipu currently lacks a formal machine-readable encoding suitable for systematic or computational analysis. This talk introduces the first dedicated MEI module for jianzipu, reviewing the notation’s historical development, analyzing its structural principles, presenting the MEI.jianzipu module and its customization, and demonstrating a web-based visualization tool through a case study.

By integrating a gesture-based East Asian logographic notation into a standardized encoding framework, this work establishes a reference model for future encodings of character-based systems, including gongchepu in China; kinko-ryū shakuhachi, koto kinpu, and Okinawan kunkunshi notations in Japan; and the Geomungo Hapjabo notation in Korea. As the first MEI module centered on gesture rather than pitch, it may also offer insights for representing gesture-centric practices in contemporary Digital Musical Instrument (DMI) performance.

Co-presented by Yifan Huang and Eto Sun.

Zih-Syuan Lin — *What Counts as a Correct Score? Musical Assumptions in Transcription Evaluation*

What counts as a correct score when music is represented as data? In piano music transcription, recovering pitches and timings is only part of the transition from a performance to a score that musicians can read and use. Decisions about musical structure, notation, and visual organization raise further questions about what an evaluation should capture.

Drawing on a review of piano transcription and related work on music notation representations, this presentation examines how evaluation embodies assumptions about musical knowledge. What an evaluation measures depends on how music is represented, what is used as a

reference, and which differences are counted as errors. Agreement with a reference therefore needs to be understood in relation to the task and its intended use.

I consider the relationships among event “fidelity,” notational structure, visual organization, and human use, asking how these perspectives change the meaning of a “correct” output. This ongoing conceptual work aims to clarify the scope of existing evaluation approaches and identify questions that require additional forms of evidence. It contributes to a broader discussion of how computational approaches formalize musical knowledge through the criteria by which their outputs are judged.

Hanwen Zhang — *A Context-Aware Piano Performance Encoding Framework for Sustain Pedaling*

Although the sustain pedal is fundamental to classical piano performance, pedaling remains poorly documented. Conventional printed editions provide only skeletal cues, typically indicating pedal depression and release without conveying depth, motion, or subtle timing. In contrast, modern reproducing pianos make high-resolution pedal data widely accessible. Pedal depth values for thousands of performances are recorded and stored as Control Change 64 (CC64) in performance MIDI files. Recent work has demonstrated pedal-depth estimation from audio recordings. However, current music encoding formats and performance-extension frameworks either ignore pedal information or include only raw depth data; none provide musically meaningful representations for understanding the musical intent behind pedal movement. This limitation prevents researchers from addressing fundamental questions about pedaling style, such as comparing the “wetness” of different interpretations or characterizing coloristic pedal effects.

To address this gap, we propose a three-layer context-aware encoding framework that captures pedaling as relational gestures shaped by rhythmic and harmonic context. The physical layer preserves pedal depth trajectories. The contextual layer aligns physical events with score structure through relative pedal timing and articulation ratio. The semantic layer abstracts these descriptors into interpretable labels such as rhythmic pedaling, harmonic pedaling, coloristic blur, and anticipatory pedaling. The framework enables systematic comparison of pedaling styles across performances and supports MIR tasks including piano transcription, expressive performance rendering, and computational music analysis.

Yashar Talebirad — *What Makes Sound Layers Clash? Human Judgments, Signal Features, and Audio Models*

Two recordings can sound fine apart and wrong together. A system that layers musical and environmental recordings has to choose which combinations work, but what does it mean for sounds to clash?

Using examples from CoSound, a prototype for shared soundscapes that adapt to listener feedback, this talk compares human listening with digital signal processing (DSP) and audio-capable large language models. An LLM may identify the sources yet miss how they interact; a DSP measure may flag tension in a mix that listeners accept. A negative reaction might come from conflicting pitches, a piercing sound in one layer, or a mix that is simply too busy.

A single “compatible/incompatible” label folds these different judgments into one. This talk examines what each approach can tell us, including whether an LLM’s answer depends on the audio it was given. The longer-term goal is a benchmark for how models hear relationships between simultaneous sounds, while keeping acoustic measurements and listener judgments distinct.

Wenye Ma — *What Matters for Music-Centered Recognition in Audio-Language Models?*

Audio-language models (AudioLMs) provide a flexible natural-language interface for audio understanding, but it remains unclear how well they represent and preserve task-relevant musical information for structured recognition tasks. This talk examines how musical knowledge is encoded, transformed, and ultimately mapped to labels within AudioLM pipelines.

This work studies four recognition tasks—instrument recognition, key detection, genre tagging, and environmental sound recognition—through three complementary experiments. First, representations are probed at the audio encoder, connector, and LLM audio-token stages of existing AudioLMs to determine where task-relevant information remains accessible. Second, using controlled models with the same frozen audio encoder, direct classification is compared with LLM-based label generation. Third, the talk investigates how encoder adaptation, connector pretraining, and multi-encoder fusion affect joint recognition performance.

The results show that recognition performance cannot be explained by simple information loss through the AudioLM pipeline. Task-relevant information often remains available beyond the encoder, while the language interface changes recognition behavior in a strongly task-dependent manner. These findings highlight how modeling choices shape the way musical information is represented, combined, and translated into computational labels.