September
21-23, 2026

4th International Conference on Computational and Cognitive Musicology

Julius-Maximilians-
UNIVERSITÄT WÜRZBURG

VOGEL Stiftung

zpd

Universitätsbund Würzburg

https://go.uniwue.de/icccm26

The open-source LaTeX template, AMCOS_booklet, used to generate this booklet is available at `https://github.com/maximelucas/AMCOS_booklet`

Contents

About

Welcome

to the Fourth International Conference on Computational and Cognitive Musicology, welcome to Würzburg!

When we decided to host ICCCM this year, we did not expect that we would receive such a great number of submissions from almost all corners of the world. The program comprises a truly broad and rich spectrum of topics, research questions, and methodologies, centered around approaches situated within computational and/or cognitive musicology. This perfectly fits this year's conference theme:

"Studying Music Across Time, Space, and Cultures,"

and shows that computational and cognitive approaches can be relevant to virtually all areas of music research.

The conference takes place in the Zentrum für Philologie und Digitalität (Center for Philology and Digitality, ZPD), a central institution of the Würzburg University and a place for research and interdisciplinary exchange for the humanities and digital methodologies. This building was opened in 2023, and we believe that we are the largest event happening here so far.

With 115 registered participants from 23 countries on 6 continents—16 oral presentations and 66 posters—this is probably the largest ICCCM yet. This makes us very hopeful for the future, both for the conference itself, but also for our diverse fields of research: more than half of all attendants are PhD students or at earlier career stages.

Tomorrow, composer and professor for artistic research Artemi-Maria Gioti (Mozarteum University Salzburg, Austria) will give her invited keynote on "Music Composition as Critical Data Practice," which will enrich our event with perspectives based in artistic practice and critical reflection. We are already looking forward to inspiring discussions.

We are grateful to everyone who helped realize this conference through financial support: Universitätsbund Würzburg e.V., Vogel Stiftung Dr. Eckernkamp, as well as ZPD and the Institute for Music Research of Julius-Maximilians-Universität Würzburg. Among other things this allowed us to award 7 travel grants for international participants. Finally, we thank the Steering Committee of the Digital Musicology Study Group of the International Musicological Society for their confidence in us as organizers.

We are very much looking forward to these coming days, to interesting discussions and inspiring exchanges.

Würzburg, September 21, 2026
Fabian C. Moss and Christof Weiß
Conference Chairs

Program

T: Talk, K: Keynote.

Monday, September 21, 2026

13:15–13:30	**Opening and Welcome**		
13:30–15:00	**Oral 1: Folk Music Across Cultures** Session Chair: Christof Weiß		
13:30–13:50	T	**Stephen Ayesu** University of Education, Winneba, Ghana	Assessing the Archival Readiness of Ghanaian Traditional Music Collections for Computational Ethnomusicology
13:50–14:10	T	**Lisette Ant; Anna Aljanaki** University of Tartu; University of Music and Performing Arts Graz	Vocal Cloning as a Method for Cross-Cultural Vocal Timbre Analysis: A Case Study on Finno-Ugric Folk Song
14:10–14:30	T	**Eyob Musie Demas** Mekelle University	Computational Approaches to Cultural Symbolism and National Identity in Traditional Tigrigna Songs of Tigray
14:30–14:50	T	**John McBride; Tecumseh Fitch; Amanda LeRoi** Institute for Basic Science; University of Vienna; Imperial College London	Evolution and Ecology of Irish Folk Tunes
15:00–15:30	**Coffee Break**		
15:30–17:00	**Poster Session A**		
17:00–18:00	**Oral 2: Harmony and Tonality** Session Chair: Claire Arthur		
17:00–17:20	T	**Mirjam Visscher; Peter Van Kranenburg; Frans Wiering** Utrecht University; Utrecht University, Meertens Institute; Utrecht University	300 Years of Pitch Content: The Transition from Modality to Tonality 1450–1750
17:20–17:40	T	**Noam Rotem; Lior Abel** The Open University of Israel; Ben Gurion University of the Negev	Compressibility of Chord-Progression Corpora as a Measurement of a Musical Dialect's Harmonic Language

17:40–18:00	T	**Uri Rom; Michael Savin; Martin Rohrmeier** Tel Aviv University; Tel Aviv University; EPFL	Harmonic Root Progressions as a Representation of Composition Style and Period
19:00–	**Jam Session**		

Tuesday, September 22, 2026

09:30–10:30	**Oral 3: Music Perception & Cognition** Session Chair: Joshua Albrecht		
09:30–09:50	T	**Rasmus Levin; Ram Reuven** Norwegian Academy of Music; Norwegian Academy of Music	Assessing Musical Expectation in Music-Theoretical and Music-Cognition Corpora: Toward a Memory-Systems Framework
09:50–10:10	T	**Markus Foramitti; Christoph Reuter; Mauricio Martins; Tudor Popescu** University of Vienna; University of Vienna; University of Vienna; University of Vienna	Do Audio Features of Popular Music Mirror Lyrical Trends in Stress, Valence, and Complexity Over Five Decades?
10:10–10:30	T	**Danbinaerin Han; Dasaem Jeong; Juhan Kam** KAIST; Sogang University; KAIST	Where in the Sound Does Social Function Shape Musical Form? A Layered Audio Analysis of Korean Folk Songs
10:30–11:00	**Conference Group Photo**		
11:00–11:30	**Coffee Break**		
11:30–12:30	**Poster Session B**		
12:30–14:00	**Lunch Break**		
14:00–15:00	**Oral 4: Melody** Session Chair: Lele Liu		
14:00–14:20	T	**Andres Puente; Jan Hajič, jr.; Gustavo Ballen** Pontifical Xavierian University; Charles University; Universidade Estadual Paulista	Modeling Gregorian Melodic Diversity With Biogeography and Comparative Phylogenetics
14:20–14:40	T	**Polina Proutskova et al.** Queen Mary University of London et al.	VocalNotes: Cross-Cultural Perspectives on Musical Note Perception Through Comparative Transcription
14:40–15:00	T	**Claire Arthur** Georgia Institute of Technology	Stuck in the Loop: A Computational Evaluation of Self-Cueing Mechanisms in an Earworm Corpus
15:00–15:30	**Coffee Break**		
15:30–17:00	**Poster Session C**		
17:00–18:00	K	**Artemi-Maria Gioti** Mozarteum Salzburg, Austria	Music Composition as Critical Data Practice
19:00	**Conference Dinner**		

Wednesday, September 23, 2026

10:00–11:00	**Oral 5: Corpora and Archives** Session Chair: Johannes Hentschel		
10:00–10:20	T	**Katelyn Emerson; Peter Harrison** University of Cambridge; University of Cambridge	Performers in Context: A New Multimodal Dataset for Computational and Cognitive Studies of Organists' Interpretations
10:20–10:40	T	**Maik Köster; Markus Neuwirth** Anton Bruckner University; Anton Bruckner University	A Multivalent Corpus Study of Nineteenth-Century String Quartet Expositions
10:40–11:00	T	**Stefan Balke; Manuel Peters; Meinard Müller** International Audio Laboratories Erlangen; International Audio Laboratories Erlangen; International Audio Laboratories Erlangen	Wind Music as a Research Scenario for MIR and Computational Musicology
11:00–11:30	**Coffee Break**		
11:30–13:00	**Poster Session D**		
13:00–13:15	**Closing Remarks**		

Keynote by Artemi-Maria Gioti

Music Composition as Critical Data Practice

Artemi-Maria Gioti (Mozarteum Salzburg)

This talk will draw on autoethnographic materials from the speaker's compositional practice to examine the critical insights that artistic research in composition can offer into machine learning and data practices. It will focus specifically on the perspectives that emerge through the creation of datasets, the training of machine learning models, and their deployment in live performance settings. Bringing an autoethnographic perspective on data-making practices into dialogue with theoretical frameworks from critical data studies, it will propose a deconstructive critique of data as material, processual, and relational, foregrounding the aesthetic decisions embedded in processes of datafication. Finally, it will explore how errors and failures in machine learning systems can function as sites of critical inquiry into the epistemological assumptions that underlie machine learning.

Artemi-Maria Gioti (`https://www.artemigioti.com/`) is a composer and artistic researcher conducting critical research at the intersection of music and artificial intelligence (AI). She is Professor of Artistic Research in Music at Mozarteum University Salzburg. Her compositional work focuses on interactive works involving reciprocal, real-time interaction between human performers and computer music systems incorporating machine learning (ML). Through this practice, she explores how technology reconfigures musical practices and ontologies, such as authorship and the musical work. Her research brings autoethnographies of the compositional process into dialogue with theoretical frameworks from critical AI and critical data studies, to investigate questions of data materiality, data semiotics and ML epistemology. Gioti is a core team member of the ERC-funded project MusAI (Music and Artificial Intelligence: Building Critical Interdisciplinary Studies), led by Georgina Born. She previously served as the Principal Investigator of the FWF-funded research project Inter_agency. She holds a doctoral degree in Music Composition from the University of Music and Performing Arts Graz with previous studies in Composition, Electroacoustic Composition and Computer Music at the University of Macedonia (Thessaloniki, Greece), the University of Music and Performing Arts Vienna and the University of Music and Performing Arts Graz (Austria). Her previous academic appointments include positions at the University of Music "Carl Maria von Weber" Dresden (Germany), UCL's Anthropology Department (UK) and the Institute of Electronic Music and Acoustics (IEM) of the University of Music and Performing Arts Graz (Austria).

List of Abstracts – Talks

Oral 1: Folk Music Across Cultures

Assessing the Archival Readiness of Ghanaian Traditional Music Collections for Computational Ethnomusicology

Stephen Ayesu (University of Education, Winneba, Ghana)

The growing field of computational ethnomusicology has increased the demand for music collections that are not only digitally preserved but also discoverable, accessible, interoperable, and reusable for computational analysis. However, the limited availability of well-documented and computationally accessible African music collections continues to constrain the representation of African musical traditions within global computational music research. Despite decades of audio-visual documentation of Ghanaian traditional music, little is known about the extent to which existing institutional collections are prepared to support computational ethnomusicological research and integration into emerging digital heritage infrastructures. This proposed study assesses the archival readiness of Ghanaian traditional music collections housed in five repositories: the J.H. Nketia Archive, the Ghana Broadcasting Corporation Gram Library, the University of Ghana Music Department Library, the University of Education, Winneba (UEW) Music Department Library, and the UEW African Unit Digital Media Collections. The study has a single objective: to assess the readiness of these collections for computational ethnomusicology through the evaluation of metadata quality, documentation practices, technical recording quality, machine-readability, interoperability, and digital preservation standards. Employing qualitative content analysis informed by computational ethnomusicology, archival science, and digital humanities, the study will evaluate selected audio-visual recordings and associated documentation using the FAIR Data Principles and Dublin Core Metadata Standards as analytical frameworks. Particular attention will be given to metadata completeness, descriptive consistency, accessibility, interoperability, annotation depth, format sustainability, and computational usability. The study is expected to produce an Archival Readiness Assessment Framework for African traditional music repositories, providing practical guidance for enhancing metadata standardisation, computational accessibility, and future applications in music information retrieval and digital music corpus development. By advancing the computational inclusion of Ghanaian traditional music collections, the study contributes an African perspective to international discussions on digital music archives, cultural heritage preservation, and the responsible digital transformation of music collections.

Vocal Cloning as a Method for Cross-Cultural Vocal Timbre Analysis: A Case Study on Finno-Ugric Folk Song

Lisette Ant (University of Tartu); Anna Aljanaki (University of Music and Performing Arts Graz)

Timbre is among the most complex properties of musical sound, and its assessment remains largely listener-dependent and therefore subjective. In folk music, characterising vocal timbre is important for documenting and comparing the distinctive features and identities of different singing traditions [1, 2]. A central methodological obstacle to cross-cultural comparison is that the acoustic features of a sung recording are shaped simultaneously by several entangled factors: melody, the language and its phonemic content, vocal register, singing technique. Some of these factors are voluntary choices which distinguish a singing tradition, and some are confounds. This study asks whether vocal cloning can be used to disentangle timbre from the language-driven components, and thereby make cross-cultural timbral comparison more measurable [3,4]. We compiled a research corpus from the audio archives of the Estonian Folklore Archives and the Hungarian Heritage House, covering the singing traditions of five Finno-Ugric peoples (Estonians, Karelians, Khanty and Mansi, Sámi, and Hungarians). The corpus is restricted to a capella songs performed by adult female informants. After quality control, denoising, music–speech segmentation, and silence removal, the corpus comprised 2,208 ten-second clips from 110 singers. The core of the method applies a zero-shot voice conversion model [4] that transfers the timbral character of each archival recording onto a neutral source signal sung on a single vowel, removing native-language phonemic content while preserving timbral identity. Spectral features, formants, harmonics-to-noise ratio and other voice parameters were computed on both the original and the cloned material [5]. These characteristics were compared across nationality using descriptive statistics and t-SNE cluster maps. In addition, random-forest vocal tradition classification model was trained with singer-independent cross-validation. Clear between-group differences emerged across spectral features, formants, and voice quality, and these patterns persisted in the cloned material, though with reduced contrast, indicating that cloning smooths acoustic profiles without erasing culture-specific information. Notably, classification accuracy rose from 0.777 on the original material to 0.814 on the cloned material, suggesting that removing linguistic content lets the model focus on the timbral features that carry ethnically distinctive information. The results indicate that vocal cloning offers a workable, although in many ways constrained methodological framework for isolating timbre from linguistic content and supports comparative analysis of vocal timbre across cultures.

References

[1] Belin, P., Fecteau, S., Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. Trends in Cognitive Sciences, 8(3), 129–135.

[2] Gómez, E., Haro, M., Herrera, P. (2009). Music and geography: Content description of musical audio from different parts of the world. Proceedings of ISMIR, 753–758.

[3] Berrak Sisman, Junichi Yamagishi, Simon King, and Haizhou Li. 2020. An Overview of Voice Conversion and Its Challenges: From Statistical Modeling to Deep Learning. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 29 (2021), 132–157. https://doi.org/10.1109/TASLP.2020.3038524.

[4] Liu, S. (2024). Zero-shot Voice Conversion with Diffusion Transformers. arXiv:2411.09943.

[5] Jadoul, Y., Thompson, B., de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. Journal of Phonetics, 71, 1–15, https://doi.org/10.1016/j.wocn.2018.07.001.

Computational Approaches to Cultural Symbolism and National Identity in Traditional Tigrigna Songs of Tigray

Eyob Musie Demas (Mekelle University)

This paper examines the use of computational ethnomusicology in analyzing cultural symbolism and nationalist expression in traditional Tigrigna songs from Tigray, Ethiopia. Traditional Tigrigna music plays an important role in preserving cultural memory, identity, spirituality, and social values. Despite the cultural significance of these musical traditions, most previous studies have relied primarily on qualitative interpretation. This study proposes a computational and digital approach to organizing and analyzing recurring thematic patterns in selected traditional songs produced between 2008–2011 E.C. Using thematic coding and digital text organization methods, the research investigates how symbolic expressions, patriotic narratives, and traditional cultural values are represented across multiple songs and performers. The study focuses particularly on themes of homeland, identity, spirituality, love, and social values reflected in lyrical content and performance contexts. Computational approaches are used to classify and compare recurring themes, allowing for more systematic documentation and interpretation of musical materials. The paper argues that computational ethnomusicology can provide valuable tools for preserving and studying African musical heritage while maintaining the cultural meanings embedded within oral traditions. In addition, the study demonstrates how digital musicology may support the documentation of endangered musical cultures and expand the visibility of African traditional music within international academic discourse. By integrating ethnomusicological interpretation with computational thematic analysis, this research contributes to broader discussions on digital humanities, computational musicology, and the preservation of indigenous musical knowledge systems.

Evolution and Ecology of Irish Folk Tunes

John McBride (Institute for Basic Science); Tecumseh Fitch (University of Vienna); Armand LeRoi (Imperial College London)

Folk tunes are evolving melodic sequences that undergo change, and live or die in an ecosystem of performers and listeners. In this work I use tools from evolutionary biology and ecology to study two questions: How do tunes change? What makes some more successful than others? The study draws on a remarkable dataset from TheSession.org, where users have posted 50,000 variants of (mainly Irish) folk tunes. Users add tunes to 'tunebooks', giving us a metric of popularity change over 14 years. In the first part, I apply bioinformatics sequence-alignment tools to tune variants and highlight the need for new rhythm-aware alignment methods. I solved the alignment problem for tunes which have rigid structures (e.g, Irish dance tunes), enabling automated alignment of 80,000 pairs of melodic sequences. The analysis reveals several constraints on tune evolution: pitch stability correlates with prevalence, note substitutions mirror and preserve existing melodic motion, metric stability predicts change, and repetition patterns are conserved across variants. These findings lead to testable hypotheses about tune evolution that can be probed in the lab. In the second part, I analyse tune popularity trajectories using models from ecology. I find no strong frequency-dependent selection, but I do find per-tune selection: some tunes grow at rates faster or slower than a neutral model predicts. A machine-learning model explains 25% of the variance in tune selection coefficients. Most of the explained variance is attributable to recording prestige (e.g., Spotify popularity), and a smaller share to intrinsic melodic properties. Surprisingly, the simplest tunes are the fittest. Measuring time-variant selection coefficients lets me pinpoint precise moments where a recording boosted tune fitness. These precise estimates of tune success invite the tantalizing prospect of describing what makes a tune inherently catchy.

Oral 2: Harmony and Tonality

300 Years of Pitch Content: The Transition from Modality to Tonality 1450–1750

Mirjam Visscher (Utrecht University); Peter van Kranenburg (Utrecht University; Meertens Institute); Frans Wiering (Utrecht University)

The period between 1450 and 1750 marks the birth of tonality. Over time, the pitch system shifted from eight or twelve modes, depending on the music theorist you are consulting, to two keys. This shift is characterised by many musical developments, such as instrumentation, development of counterpoint and harmony, the rise of transposition, and chromatisation of pitch content. This study investigates the development of pitch content using the CANTOSTREAM corpus, containing 8,642 compositions. Pitch class profiles relative to the final serve as the main feature, complemented with an inspection of the development of tonal complexity. First, reference pitch class profiles for the modal families, and for the major and minor keys are created, using k-means clustering as the main technique. 200 compositions out of the corpus are set apart to create the major and minor reference profile. For creating the modal reference profiles, the CANTO-JRP corpus, based on the Josquin Research Project, is used. Then, the remaining 8,442 compositions are assigned one mode or key label. These labels are then analysed over time. Finally, the lydian mode proved to be undetectable; an analysis why this is the case is added using features from the Polymode corpus. The main contributions of this study are reference pitch class profiles for modes and keys, an initial historical analysis of the development of modes and keys, and a case study analysing the lydian mode.

Compressibility of Chord-Progression Corpora as a Measurement of a Musical Dialect's Harmonic Language

Noam Rotem (The Open University of Israel); Lior Arbel (Ben Gurion University of the Negev)

Different cultures, eras, genres, and composers favor different harmonic vocabularies; each constitutes a musical dialect with its own statistical regularities. Existing computational analysis of such dialects uses chromaticism counts, Markov-chain models of chord transitions, and curated pattern matching. Dictionary compression has been used as a tokenization step in symbolic-music modeling, but not as a per-corpus density measurement. Compression efficiency measures how much regularity a language contains: predictable, formulaic dialects compress tightly; varied ones do not. Bits-per-symbol of an optimal encoding integrates vocabulary breadth, repetition, recurring-pattern length, and inter-song shared vocabulary into a single measurement: the dialect's information density. This principle, well-established for natural language, is here applied to musical dialects, treating each song's chord progression as a text in its dialect's language. A Byte-Pair Encoding (BPE) dictionary, a standard dictionary-compression algorithm, is trained on each corpus's key-normalized chord progressions by iteratively merging the most-frequent adjacent pair into multi-chord patterns of growing length. The Shannon entropy of the token sequence, expressed in bits per original chord position, is the resulting information-density measurement. Israeli popular music differs from US along several cultural axes (geographic origin, constituent traditions, and historical age), making it a useful cross-cultural case. The analysis compares three preference-based corpora at matched n=450: McGill Billboard (US chart hits 1960-1991); Israeli singalong songs that were also Israeli chart hits 1968-1991; and a harmonically-dense Israeli sub-collection as sensitivity control. Measured information densities: McGill 1.95 bits/chord; Israeli chart hits 2.36 (Cohen's d = 0.64 vs McGill); sensitivity control 2.92 (d = 1.37 vs McGill). The discovered codebook reveals corpus-specific patterns: US top patterns are major-key vamps such as F-C or Am-F-G-C; Israeli chart hits feature harmonic-minor cadences like Dm-E-Am and Mediterranean modal cadences like F-G-Am, absent from US; the sensitivity control features extended jazz-style cadences such as Bm7b5-E-Am. This case study demonstrates the potential of compression-based methods to measure, compare, and further analyze the harmonic information density of musical-dialect corpora.

Harmonic Root Progressions as a Representation of Composition Style and Period

Uri Rom (Tel Aviv University); Michael Savin (Tel Aviv University); Martin Rohrmeier (EPFL)

Since Rameau's introduction of the concept of harmonic roots [1], root progressions—that is, the intervals between adjacent chord roots—have become an enduring object of music-theoretical inquiry [2–5]. Although they suppress such harmonic features as chord quality, inversion, and scale-degree context, root progressions make it possible to compare abstract patterns of harmonic motion across repertoires. Recent research [6–8] suggests that their frequency patterns vary across style periods and individual composers. Unlike approaches tied to local key context, we model root progressions across changing tonal regions, treating, for example, falling-fifth sequences as equivalent whether they remain within a key, include local tonicizations, or modulate. We engage the Distant Listening Corpus, which comprises harmonic annotations of 1,284 pieces by 36 composers from late Renaissance to early twentieth century [9]. Drawing on root information provided in the harmonic labels, we derive a vocabulary of 34 root-interval symbols from adjacent chord roots (-1 denotes a descending perfect fifth). We then examine the behavior of unigrams, bigrams, trigrams, and 4-grams, each consisting of sequences of root progressions. In preliminary analyses, we obtained per-piece vectors of unigrams through 4-grams from the annotations, weighting each n-gram by its cumulative temporal duration. After normalizing these vectors as probability distributions, we calculated pairwise Jensen-Shannon distances among them. A Gaussian mixture model at the bigram level, with six components, reveals non-random structure but no clean stylistic separation. The strongest signal is the emergence of two distinct Baroque subgroups, one of which attains a Baroque-cluster purity of 61%. In the next stage, we model root-progression patterns over time as gradual transformations while seeking to identify specific n-grams as markers of stylistic change. Using supervised learning methods (e.g., support vector machines with cross-validation), we test how well style periods and individual composers can be predicted from root-progression vectors.

References

[1] Rameau, J.-P. (1722). Traité de l'harmonie réduite à ses principes naturels: Divisé en quatre livres. De l'imprimerie de J.-B.-C. Ballard.

[2] Schenker, H. (1906). Harmonielehre (Neue musikalische Theorien und Phantasien, Vol. 1). J. G. Cotta'sche Buchhandlung Nachfolger.

[3] Schönberg, A. (1911). Harmonielehre. Universal-Edition.

[4] Schoenberg, A. (1954). Structural functions of harmony. W. W. Norton & Company.

[5] Sadai, Y. (1980). Harmony in its systemic and phenomenological aspects. Yanetz.

[6] Hedges, T., & Rohrmeier, M. (2011). Exploring Rameau and beyond: A corpus study of root progression theories. In C. Agon, E. Amiot, M. Andreatta, G. Assayag, J. Bresson, & J. Mandereau (Eds.), Mathematics and Computation in Music (Lecture Notes in Computer Science, Vol. 6726, pp. 334–337). Springer.

[7] Harasim, D., Schmidt, S. E., & Rohrmeier, M. (2016). Bridging scale theory and geometrical approaches to harmony: The voice-leading duality between complementary chords. Journal of Mathematics and Music, 10(3), 193–209. https://doi.org/10.1080/17459737.2016.1216186

[8] Moss, F. C. (2019). Transitions of tonality: A model-based corpus study [Doctoral

dissertation, École Polytechnique Fédérale de Lausanne]. EPFL Infoscience. https://doi.org/10.5075/epfl-thesis-9808
[9] Hentschel, J., Rammos, Y., Neuwirth, M., & Rohrmeier, M. (2025). A corpus and a modular infrastructure for the empirical study of (an)notated music. Scientific Data, 12, Article 662. https://doi.org/10.1038/s41597-025-04976-z

Oral 3: Music Perception and Cognition

Assessing Musical Expectation in Music-Theoretical and Music-Cognition Corpora: Toward a Memory-Systems Framework

Rasmus Levin (Norwegian Academy of Music); Ram Reuven (Norwegian Academy of Music)

This paper argues that expectation in harmony textbooks used in Scandinavian higher education, including Anglo-American textbooks, is underspecified, framing expectation in terms of stylistic implications rather than as the subjective experience identified by contemporary music-cognitive research. This framing obscures the distinction between descriptions of style and listeners' experience, yielding ambiguity that undermines clarity, empirical validity, and engagement with contemporary music-cognitive research. This ambiguity encourages teachers and students to treat stylistically conditioned listening as "natural," downplaying its basis in long-term stylistic exposure. The argument unfolds in four stages, supported by a new database of 810 concepts across 37 music-theoretical and music-cognitive sources. First, a survey of 30 harmony textbooks from 1933–2025 shows that they conflate stylistic implications with listeners' experiences, inadvertently treat expectation as a single type, and rarely reflect empirical research on expectation. Second, examination of advanced analytical literature and studies linking music cognition and music theory shows that even empirically informed work has related limitations: reliance on outdated empirical findings, overlapping and compound expectation concepts, limited coverage of expectation types and conflation of style structures with schemas and expectation. Third, a review of music-cognitive research on predictive coding and expectation demonstrates conceptual overlap, due in part to imprecise terminology, weak engagement with contemporary memory models, and underdeveloped treatment of working-memory-based expectation. Fourth, the paper develops a memory-based taxonomy of principal expectation types empirically grounded in cognitive and neuroscientific models. This memory-based framework, rather than schema-based models, better explains distinct expectation types and their mediation across learning, knowledge, listening experience, and harmonic syntax while bridging lexical and subjective conceptualizations of style. The paper concludes by outlining how this framework resolves ambiguities between stylistic implications and subjective experiences, offers terminology that clarifies relations between lexical and subjective perspectives, and suggests applications to curriculum design, analytical writing, and listening pedagogy.

Do Audio Features of Popular Music Mirror Lyrical Trends in Stress, Valence, and Complexity Over Five Decades?

Markus Foramitti (University of Vienna); Christoph Reuter (University of Vienna); Mauricio Martins (University of Vienna); Tudor Popescu (University of Vienna)

A recent large-scale study (Foramitti et al., 2025) analysed over 20,000 lyrics from the US Billboard Hot 100 between 1973 and 2023, finding a significant long-term decline in lyrical complexity and positive sentiment, and an increase in stress-related language. Notably, major societal shocks such as COVID-19 coincided with attenuations rather than amplifications of these trends, consistent with emotion-incongruent music selection and collective escapism. However, it remains unknown whether these patterns are specific to the textual dimension of songs or whether (and in what form) corresponding trends are also present in the musical signal. We addressed this question using the same Billboard sample, drawing on the Million Song Dataset and pre-collected Spotify audio descriptors to obtain musical analogues to the lyrical constructs examined in Foramitti et al. Features included valence, energy, mode, tempo, and tonal tension distances as proxies for emotional content and arousal. Complexity was operationalised via LZ77 compressibility applied to melodic and harmonic sequences. To validate these database-derived measures, a stratified sample of approximately 200 songs spanning the full fifty-year period was independently processed from raw audio recordings and cross-referenced against dataset values. We present findings on the degree to which long-term trends and event-related deviations observed in lyrics are replicated in the musical domain, and discuss results in terms of Mood Management Theory. This multimodal analysis offers a more complete assessment of whether popular music, as a composite signal, functions coherently as a collective mood regulation vehicle; and specifically whether its lyrical and musical dimensions carry emotional information in congruent or in complementary ways.

References

[1] Foramitti, M., Nater, U. M., Lamm, C., & Martins, M. (2025). Societal crises disrupt long-term increases in stress, negativity, and simplicity in US Billboard song lyrics from 1973 to 2023. Scientific Reports, 15(1), 41733.
https://doi.org/10.1038/s41598-025-28327-5

Where in the Sound Does Social Function Shape Musical Form? A Layered Audio Analysis of Korean Folk Songs

Danbinaerin Han (KAIST); Dasaem Jeong (Sogang University); Juhan Nam (KAIST)

Social function is thought to shape musical form, but where in the sound this happens and how it can be measured remain open questions. In Korean tosok minyo, the orally transmitted local folk songs, each song is filed by the use it served, and the question must be answered from the recordings themselves. From 11,344 field recordings we measured form in two layers, how densely and regularly a song is cut into phrases and how much melodic material returns and how regularly, and beneath them the delivery: whether one voice sings on or a group answers, and with what vocal quality. Validated against expert judgements, continuous pitch contours captured this repertoire's melodic relations better than note-level transcription. With song length controlled, function effects appeared in the delivery and in both layers of form, largest in the delivery and in phrase density, whereas in the return of material only its phase alignment was robust, and much of the rest was mediated by song type. The same measures were examined across work types, regions, and published transcriptions. The layered protocol localises the function effect and can be applied to other oral traditions in which a unit of phrase segmentation is agreed.

Oral 4: Melody

Modeling Gregorian Melodic Diversity With Biogeography and Comparative Phylogenetics

Andres Puente (Pontifical Xavierian University); Jan Hajič, jr. (Charles University); Gustavo Ballen (Universidade Estadual Paulista)

Should we wish to examine the roots of a shared European musical identity, and at the same time its internal diversity, one can hardly find a tradition that exemplifies this balance better than Gregorian chant. It served as an expression of cultural — and political — cohesion of Latin Europe (Rankin, 2022; Hucke, 1980; Hiley, 1993, p.514-523). Substantial effort was devoted to conserving melodies exactly: the schola cantorum (Fassler, 1985), the expenses of producing liturgical manuscripts, the technology of staff notation (Rankin, 2018; Atkinson, 2008, ch.6). However, even notated, chant melody was subject to forces acting on oral tradition (Treitler, 1981; Karp, 1998; Kelly, 2023), and written transmission in manuscript culture is no guarantee of accuracy either (Roelli, 2020, ch. 4.3). The resulting melodic diversity is barely understood. Some melodic "dialects" have been proposed: the East vs. West Frankish split (Wagner, 1925; Blachly, 1990), with further differentiation observed in western sources (Froger, 1978); the Cistercian redaction (Glasenapp, 2020), or a Neustrian dialect (Goudesenne, 2019). However, each has been identified with different (and not always transparent or scalable) methods, so findings are difficult to combine into a more complete picture. Phylogenetic methods derived from processes of chant melody transmission have emerged as a computational alternative to model the melodic diversity of chant systematically and scalably. Already first experiments verified the existence of melodic "dialects" (Hajič jr. et al., 2025; Ballen et al., 2026), identifying the place and some institutions of origin as the main factors structuring Gregorian melodic variability. We take this methodological perspective further and study the spatial distribution of variability in chant melody using biogeographic parametric models, which consider alternative area occupation dynamics, and finding the best ones via model selection. We also measure the congruence between mass and office chants using co-phylogenetics, and the melodic consistency of individual proposed "clades" with phylogenetic comparative methods. This work illustrates the continued value of taking a cultural evolution perspective to study the vast tradition of Gregorian chant.

References

[1] Atkinson, C. M. (2008). The critical nexus: tone-system, mode, and notation in early medieval music. Oxford University Press.

[2] Ballen, G. A., Mühlová, K. H., and Hajič, J. (2026). What did the dove sing to Pope Gregory? Ancestral melody reconstruction in Gregorian chant using Bayesian phylogenetics. PLOS One, 21(5):e0350014.

[3] Blachly, A. (1990). Some Observations on the "Germanic" Plainchant Tradition. Current Musicology, (45-47):85– 117.

[4] Fassler, M. E. (1985). The office of the cantor in early Western monastic rules and customaries: a preliminary investigation. Early Music History, 5:29–51.

[5] Froger, D. J. (1978). The critical edition of the Roman Gradual by the monks of Solesmes. Journal of the Plainsong & Mediaeval Music Society, 1:81–97.

[6] Glasenapp, J. (2020). To Pray without Ceasing: A Diachronic History of Cistercian Chant in the Beaupré An- tiphoner (Baltimore, Walters Art Museum, W. 759–762). Columbia University.
[7] Goudesenne, J.-F. (2019). Émergences du Chant Grégorien : Les strates de la branche Neustro-Insulaire (687-930). Brepols.
[8] Hajič Jr, J., Lanz, V., and Ballen, G. A. (2025). Genome of melody: applying bioinformatics to study the evolution of Gregorian chant. Philosophical Transactions of the Royal Society B, 380(1940).
[9] Hiley, D. (1993). Western plainchant: a handbook. Clarendon Press, Oxford, United Kingdom. Hucke, H. (1980). Toward a new historical view of Gregorian chant. Journal of the American Musicological Society, 33(3):437–467.
[10] Karp, T. (1998). Aspects of orality and formularity in Gregorian chant. Northwestern University Press.
[11] Kelly, T. F. (2023). Oral and written transmission in chant. Taylor & Francis.
[12] Rankin, S. (2018). Writing sounds in carolingian Europe: The invention of musical notation, volume 15. Cambridge University Press.
[13] Rankin, S. (2022). Sounding the Word of God: Carolingian Books for Singers. University of Notre Dame Press.
[14] Roelli, P., editor (2020). Handbook of Stemmatology: History, Methodology, Digital Approaches. De Gruyter.
[15] Treitler, L. (1981). Oral, Written, and Literate Process in the Transmission of Medieval Music. Speculum, 56(3):471– 491.
[16] Wagner, P. (1925). Germanisches und Romanisches im frühmittelalterlichen Kirchengesang. In Bericht über den I. Musikwissenschaftlichen Kongreß der deutschen Musikgesellschaft in Leipzig 1925, pages 21–34, Leipzig, Germany.

VocalNotes: Cross-Cultural Perspectives on Musical Note Perception Through Comparative Transcription

Polina Proutskova (Queen Mary University of London); Gakuto Chiba (Keio University); Miranda Crowdus (Concordia University); Yulia Nikolaenko (Independent Researcher); Yuto Ozaki (Keio University); Lawrence Shuster (Cornell University); Olga Velichkina (French Society for Ethnomusicology); Zhaoxin Yu (Shandong University of Arts); Wey Yue (Shandong University of Arts); Yukun Li (Queen Mary University Of London); Andrew Killick (University of Sheffield); John McBride (Institute for Basic Science); Elizabeth Phillips (McMaster University); Patrick Savage (Keio University)

Music transcription is widely used in musicological research, yet relatively little attention has been paid to the extent to which expert transcribers agree when representing vocal performances. This question is particularly relevant in vocal traditions where the pitch curve varies continuously and does not map neatly onto discrete symbolic representations. This paper presents findings from the VocalNotes project, an international collaborative study investigating transcription disagreement in vocal music across multiple musical traditions [1]. The project brought together researchers and practitioners from five traditions—Japanese folk song, Chinese Bangzi opera, Russian village singing, Alpine yodelling, and Romaniote Jewish chant. Participants created multiple independent transcriptions of selected recordings from their own traditions using a shared workflow and annotation framework. The methodology combines independent expert transcription with shared digital tools for segmentation, pitch correction, comparison and visualisation based on Tony, Sonic Visualiser and additional custom scripts. The project has also produced the openly available VocalNotes dataset, comprising recordings, segmentations, transcriptions, and associated annotations from participating teams [2]. Through systematic qualitative analysis of disagreements, cultural, linguistic, and musical contexts were identified that tended to give rise to divergent transcription decisions within each tradition. A cross-team synthesis identified which of these contexts were relevant across multiple traditions. The resulting framework provides a systematic way of documenting and comparing transcription disagreement across vocal traditions while preserving the cultural and analytical perspectives of individual research teams. The project demonstrated that the methodology can support a range of research questions pursued by participating teams, including relationships between theoretical and performed modes [3], performer-specific variation, the influence of singing expertise on transcription practice [4], and the consequences of transcription disagreement for modal analysis. The talk will introduce the VocalNotes methodology, dataset, and project outcomes, and outline follow-up studies and future directions.

References

[1] Proutskova, Polina, Gakuto Chiba, Miranda Crowdus, Andrew Killick, Yukun Li, John M. McBride, Yulia Nikolaenko, Yuto Ozaki, Elizabeth Phillips, Patrick E. Savage, Lawrence Shuster, Olga Velichkina, Wei Yue, Zhaoxin Yu, and Gabriel Zuckerberg. 2025. "VocalNotes: Investigating the Perception of Note Pitch and Boundaries through Varying Transcriptions of Vocal Performances from Five Musical Cultures." Analytical Approaches to World Music 13 (2).

[2] Proutskova, P., McBride, J., Ozaki, Y., Chiba, G., Li, Y., Yu, Z., Yue, W., Crowdus, M., Zuckerberg, G., Velichkina, O. and Nikolaenko, Y., 2023, November. The VocalNotes Dataset. In International Society for Music Information Retrieval Conference,

Late-Breaking/Demo.
[3] Yu, Zhaoxin, Yukun Li, and Wei Yue. 2025. “A Computational Analysis of Melodic Characteristics in Hebei Bangzi Opera: A Case Study of the ‘Sha Miao’ Aria.” Analytical Approaches to World Music 13 (2).
[4] Chiba, Gakuto, Yuto Ozaki, and Patrick E. Savage. 2025. “What Is a ‘Note’? Agreement and Disagreement in Transcriptions of Japanese Folk Songs.” Analytical Approaches to World Music 13 (2).

Stuck in the Loop: A Computational Evaluation of Self-Cueing Mechanisms in an Earworm Corpus

Claire Arthur (Georgia Institute of Technology)

Earworms, a type of (mostly) involuntary musical imagery (INMI), are frequently thought of as mere musical curiosities. Recently, however, scholars have suggested that earworms may hold valuable clues to understanding the unique nature of music and memory, studying the phenomenon through the lens of contemporary theories of cognition and memory (Arthur, 2023; Schubert, 2023). Arthur (2023) notes that most studies focus on earworm symptoms rather than the underlying mechanisms, raising questions about repetition structure and involuntary rehearsal, while Schubert (2023) highlights our limited understanding of how musical episodes organize repetition. This paper responds by proposing and computationally testing two self-cueing mechanisms for the looping effect of earworms using a novel, ecologically-valid earworm corpus. The corpus originated from an experience sampling methodology (ESM) study (Evans & Davidenko, 2022) where participants recorded spontaneously-occurring earworms over two weeks resulting in 139 unique earworm episodes. These melodic fragments were encoded symbolically within their full song context and are compared to a control corpus of pop melodies . Using sliding windows (2-measure windows, 1-measure slide), we compute two cue metrics at each window position. The first mechanism, expectancy-based cueing, posits that learned structural expectations drive looping; i.e., when a listener strongly anticipates the continuation of material. Expectancy-based cueing is measured as a weighted blend of schematic and veridical prediction probabilities computed via a "prediction by partial match" (PPM) algorithm, where schematic expectations are derived from a training corpus and veridical expectations from prior material within the same song. The second mechanism, match-based cueing, involves memory retrieval where re-encountering familiar material triggers 'backward jumps.' It is measured using Smith-Waterman sequence alignment on melodic interval and rhythm pairs against all preceding windows. For each mechanism we test whether (a) earworm songs show elevated cue affordances relative to controls, and (b) the reported earworm trigger window scores higher than non-trigger windows within the same song — offering a fine-grained account of how musical memory operates spontaneously in naturalistic contexts.

Oral 5: Corpora and Archives

Performers in Context: A New Multimodal Dataset for Computational and Cognitive Studies of Organists' Interpretations

Katelyn Emerson (University of Cambridge); Peter Harrison (University of Cambridge)

Recent research in performance science, empirical musicology, and computational music analysis has shown that musicians' interpretations are shaped by interplay between such factors as instrument characteristics, performers' attributes, and cultural context (see e.g., Palmer, 1997; Cook, 2014; Doğantan-Dack, 2021; Nijs et al., 2013; de Souza, 2017). However, most open-access multimodal performance datasets focus on audio or MIDI collected in laboratory settings, detached from information about the musicians and recording contexts. This obscures performers' backgrounds, agency, and experiences, limiting cross-disciplinary study of expressive performances (Emerson & Harrison, 2025; Maestre et al., 2017). This presentation introduces the ORgAN Dataset (Organists, Performances, And iNterviews), an open-access multimodal dataset designed to reconnect musical output with performer, instrument, and context. The ORgAN Dataset comprises audio, multi-angle video, and MIDI recordings of 30 expert organists from 10 countries performing the same two compositions on two organs with different actions in unique acoustics, totalling 28.78 hours of performances. Alongside performance data, the ORgAN Dataset provides performer metadata, including age, gender, birth country, native language, education, professional background, and anthropometric measurements (hand size, height, shoe size, and bench position). It also contains interview transcriptions of participants discussing their interpretative intentions, performance and data collection experiences, and perspectives on musical expressivity (61.9 hours). Linguistic samples include participants' readings of the same English prose and poetry, plus native-language prose from non-native English speakers. The ORgAN Dataset combines performance, contextual, biographical, interview, anthropometric, and speech data to support rich computational and cognitive investigations. These include cross-modal studies of gesture and audio, development and testing of expressive timing and articulation models, performance style classification within and between performers, and investigation into the influence of physical and biographical factors as well as the relationships between performers' intentions and realised outcomes, opening new avenues for interdisciplinary research in music, cognition, and linguistics.

References

[1] Cook, N. (2014). Beyond reproduction: Semiotic perspectives on musical performance. Musikologija, 16, 15–30. https://doi.org/10.2298/MUZ1416015C.

[2] Doğantan-Dack, M. (2021). Senses and sensibility: The performer's intentions between the page and the stage. Rivista di Analisi e Teoria Musicale, 27(1), 23–67.

[3] Emerson, K., Harrison, P.M.C. (2025). Multimodal Datasets for Studying Expert Performances of Musical Scores. Transactions of the International Society for Music Information Retrieval 8(1), 400–428. https://doi.org/10.5334/tismir.230.

[4] Maestre, E., Papiotis, P., Marchini, M., Llimona, Q., Mayor, O., Pérez, A. (2017). Enriched multimodal representations of music performances: Online access and visualization. IEEE MultiMedia 24(1), 24-34. https://doi.org/10.1109/MMUL.2017.3.

[5] Nijs, L., Lesaffre, M., and Leman, M. (2013). The musical instrument as a natural extension of the musician. In M. Castellengo, H. Genevois, and J.-M. Bardez (Eds.), Music and its Instruments (pp. 467–484). Editions Delatour France.
[6] Palmer, C. (1997). Music performance. Annual Review of Psychology, 48, 115–138. https://doi.org/10.1146/annurev.psych.48.1.115.
[7] de Souza, J. (2017). Music at hand: Instruments, bodies, and cognition. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190271114.001.0001.

A Multivalent Corpus Study of Nineteenth-Century String Quartet Expositions

Maik Köster (Anton Bruckner University); Markus Neuwirth (Anton Bruckner University)

Though they embrace a more empirical and historically sensitive approach to form, the major textbooks of the New Formenlehre (Caplin 1999; Hepokoski & Darcy 2006) remain closely married to the canonised repertoire of Haydn, Mozart, and Beethoven – thus falling short of their promised potential (Wingfield 2008). If form is understood not as the manifestation of ideal types but as a dynamic interaction between works and listener expectations emerging from experience with a repertoire, corpus studies become an important avenue for building robust models of form and grounding claims about diachronic change. To realise this potential for at least a specific formal section and genre, this paper presents preliminary evaluations on an annotated corpus (n=81) of string quartet exposition sections. The corpus selection was informed by historical repertoire studies rather than pragmatic or anachronistic criteria, which often introduce undesirable biases. (London 2022) About half of the corpus was encoded by hand from historical parts and comprises neglected composers never before included in computational research. Beyond testing this historically grounded approach, we model form as the multivalent interplay of distinct musical domains (Webster 2009; Yust 2018), represented through a combination of human annotation and automated retrieval. We developed an annotation procedure of detailed hierarchical annotations primarily build around the TimeLineAnnotator app (Martins & Gotham 2023), using a controlled vocabulary of formal functions, types and semantic modifiers, and supplement this holistic annotation with domain-specific reductions that are independent from each other: tonal trajectories, motivic relations and textural features. Analysing correlations between these sub-domains and the formal annotations, we expect even this small corpus to reveal a wide variety of compositional designs, likely to be patterernd across composers or diachronic lines. The talk lays methodological groundwork – and invites feedback – for extending this initial version of the corpus toward a richer model of nineteenth-century formal practice.

References

[1] Caplin, W. E. (1999). Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven. New York/Oxford.
[2] Hepokoski, J. & Darcy, W. (2006). Elements of Sonata Theory: Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata. Oxford/New York.
[3] Wingfield, P. (2008). "Beyond 'Norms and Deformations': Towards a Theory of Sonata Form as Reception History." Music Analysis 27(1), 137–177.
[4] London, J. (2022). A Bevy of Biases: How Music Theory's Methodological Problems Hinder Diversity, Equity, and Inclusion. In Music Theory Online 28/1.
[5] Webster, J. (2009). Formenlehre in Theory and Practice. In P. Bergé (Ed.), Musical Form, Forms & Formenlehre: Three Methodological Reflections (pp. 123–139). Leuven University Press.
[6] Yust, J. (2018). Organized Time: Rhythm, Tonality, and Form. Oxford University Press.
[7] Martins, F., & Gotham, M. (2023). "TiLiA": A timeline annotator for all. Interactive Technologies for Analysing and Visualizing Musical Structure, Copenhagen, Denmark.

Wind Music as a Research Scenario for MIR and Computational Musicology

Stefan Balke (International Audio Laboratories Erlangen); Manuel Peters (International Audio Laboratories Erlangen); Meinard Müller (International Audio Laboratories Erlangen)

Concert band and wind music represent a culturally important yet strongly underrepresented scenario in Music Information Retrieval (MIR) and computational musicology. In many countries, amateur wind ensembles play a central role in musical life, education, and community culture. At the same time, wind orchestra settings pose highly challenging research problems due to dense polyphonic textures, overlapping instrument timbres, strong articulation effects, intonation interactions, and complex room acoustics. In this contribution, we present ongoing efforts to establish wind music as a research setting connecting MIR, computational musicology, music education, and amateur music practice. As a first step, we introduced ChoraleBricks, a modular multitrack dataset based on four-part wind chorales that enables controlled studies on transcription, source separation, synchronization, and ensemble interaction. Building on this work, we now move toward full wind ensemble settings through a project centered on Gustav Holst's First Suite and Second Suite. For this project, we assembled a wind ensemble under the direction of a professional conductor. The rehearsal and concert preparation process itself becomes part of the research scenario. Reference recordings will serve as the basis for future isolated multitrack recordings following the ChoraleBricks concept. Rehearsals further provide opportunities to study ensemble interaction, acoustics, intonation, and rehearsal dynamics in realistic musical settings. Beyond dataset creation, the project explores how research infrastructures can connect musical practice, education, and community engagement. Musicians are introduced to current developments in MIR and AI-supported music analysis, while researchers gain insight into ensemble workflows and musical communication. We are also developing a web-based interface to support data access, annotation workflows, educational activities, and future community contributions. More broadly, we envision a long-term center for wind music research that fosters exchange between researchers, musicians, educators, conductors, and amateur music communities while creating new opportunities for culturally grounded and community-driven computational music research.

List of Abstracts – Posters

Poster Session A

ClassicPitch: Robust Audio-to-MIDI Transcription for Musicology

Yannik Venohr (University of Würzburg); Christof Weiß (University of Würzburg)

Audio recordings provide a promising direction for large-scale corpus studies in musicology. Compared to symbolic scores, which are often tedious to obtain, audio recordings scale much more easily, even for music where detailed scores exist in principle. Furthermore, they allow researchers to analyze performance characteristics and to consider scenarios such as improvised, orally transmitted, or electronically produced music that exists only as audio. To enable these studies at scale, there is a need for robust, efficient, and easy-to-use Automatic Music Transcription (AMT) systems that extract pitch and timing information from audio. Driven by advances in deep learning and the availability of large training datasets, AMT has improved substantially. However, most methods remain difficult to use and, more importantly, still lack robustness: AMT models that transcribe high-quality piano recordings well often fail on noisy historical ones or struggle with challenging instrumentations such as choir or orchestra. This limits their applicability across diverse repertoires and can even lead to distorted conclusions. To address this, we recently introduced ClassicPitch, a novel tool designed to robustly transcribe audio recordings to MIDI without instrument labels, enabling comparison across a wide diversity of musical works. While we have thoroughly evaluated ClassicPitch for Western classical music, we have also observed promising generalization beyond this repertoire. In this contribution, we present the tool's usage, discuss its generalization capabilities and limitations, and invite conference attendees to try ClassicPitch on their own audio examples.

Automatic Chord Recognition for Western Classical Music

Yiwei Ding (University of Würzburg); Christof Weiß (University of Würzburg)

Most corpus-level harmonic analyses of Western classical music are based on symbolic representations or entirely based on human-annotated harmony labels. This comes with limitations since the annotation process is still time-consuming and therefore infeasible on a large scale. Recent advances in Automatic Chord Recognition (ACR) allow for automatically extracting chord labels, which allows us to conduct large-scale corpus study without manually annotating harmony labels. In this contribution, we present ACR tools for Western classical music in the common-practice period. These systems can flexibly handle different input representations: First, they can take symbolic music representations such as MusicXML as input, which we refer to as symbolic ACR. This pipeline facilitates automatic chord labeling from scores. Second, they can take audio recordings as input, making use of Automatic Music Transcription (AMT) that allows us to reliably extract pitch representations from audio. The same chord recognition model as in symbolic ACR can then be used to predict chords from the extracted pitch representations. This pipeline, which we refer to as audio ACR, further facilitates audio-based large-scale corpus analysis. We test these tools on a heterogenous dataset covering a wide range of Western classical music, including different instrumentations, genres, composers, and style periods. We illustrate the effectiveness of both symbolic ACR and audio ACR tools by showing their good performance on predicting constituent pitch classes of chords. Based on the tools, we point out several potential musicological applications that analyze harmony of Western classical music on a large scale.

A Three-Century Distant Listening Study of Chromaticism and Dissonance in Western Music

Xinyi Guan (EPFL); Johannes Hentschel (Anton Bruckner University); Fabian C. Moss (University of Würzburg); Martin Rohrmeier (EPFL)

The historical narrative of increasing chromaticism and dissonance in Western music, while widely accepted, has predominantly relied on traditional close-reading analyses of individual works rather than comprehensive empirical validation. This study presents a computational analysis of a large-scale dataset—comprising over 230,000 expert-annotated harmonic labels from 1,280 musical works spanning three centuries—to quantify and trace the diachronic development and co-evolution of chromaticism and dissonance. We introduce novel quantitative indices measuring chromaticism and dissonance to evaluate stylistic evolution in different musical periods. Our findings empirically validate the established narrative of a gradual rise in both chromaticism and dissonance, and further reveal a more nuanced trajectory characterized by modal differences and a persistent diatonic foundation alongside chromatic and dissonant experimentation. Adopting a "distant listening" approach, this study provides a systematic and quantitative perspective on the evolution of chromaticism and dissonance, complementing traditional musicological approaches and providing novel, data-driven insights into the evolution of Western tonal music.

harMo13 – Computational Analysis of 13th Century Motets

Robert Eisinger (University of Cologne); Philippa Ovenden (University of Cologne); David Ubber (University of Cologne); Xuhong Qiu (University of Cologne); Frank Hentschel (University of Cologne)

harMo13 is a DFG-funded research project at the University of Cologne dedicated to the corpus-based computational analysis of 13th-century motets. One of the main goals of the project is the development of a Python-based software infrastructure to investigate the "harmonic" features of this music and to search for their consistent "grammatical" description. The data model layer handles peripheral functionalities including score input and output, parsing and rendering based on music21. Furthermore, it provides the infrastructure for representing musical events – vertically combined notes and rests at discrete time points called offsets – and interfaces for processing the corpus at different levels of abstraction (whole piece, sections, phrases, sequences of sonorities, single sonorities). An object-oriented and modular approach in the implementation using Pydantic ensures not only solid validation, but also an adequate representation of the structural elements of motets. The analytical layer treats domain-specific properties of motets as data structures, most notably the tenor cursus: a series of notes and durations in the lowest voice. Multiple repetitions of a tenor cursus are a defining structural feature of the motets in our corpus. These data structures form the basis of analytical tools, which can, for instance, identify similarities across sonorities and tenor cursus or determine musical segmentation in motets. Despite its focus on motets, the project aims to ensure reusability by separating data models and specific analytical requirements in our implementation. Among the core reusable components at this point are an adaptive categorisation system to classify sonorities, which can be applied to other musical genres (Eisinger et al. 2025), and several mechanisms to process and compare structural properties of sonorities and sequences thereof (including musica ficta variants). The presentation outlines the technical architecture of our software, discusses its current implementation status, and demonstrates selected analytical features.

References

[1] Cuthbert, Michael Scott. "music21 Documentation." music21, `https://www.music21.org/music21docs/` (accessed on 19 May 2026)

[2] Eisinger, Robert, Philippa Ovenden, Xuhong Qiu, David Ubber und Frank Hentschel. 2025. "An Explorative Approach to the Analysis of Motets, a Thirteenth-Century Musical Genre." In Proceedings of Summer School on Service-Oriented Computing 2025 (Communications in Computer and Information Science 2602), hg. von Marco Aiello, Johanna Barzen, Schahram Dustdar und Frank Leymann, 160–174. Cham, Switzerland: Springer. `https://doi.org/10.1007/978-3-032-07313-6_10`.

[3] harMo13 – Die 'Harmonik' in drei- und vierstimmigen Motetten des 13. Jahrhunderts. Eine Korpusanalyse. `https://musikwissenschaft.phil-fak.uni-koeln.de/forschungsprojekte-und-archive/laufende-projekte/harmo13-die-harmonik-in-drei-und-vierstimmigen-motetten-des-13-jahrhunderts-eine-korpusanalyse` (accessed on 19 May 2026)

[4] Pydantic Documentation, v2.13.4, `https://pydantic.dev/docs` (accessed on 19 May 2026)

A Graph-Theoretic Approach to Automatic Schema Labeling in Symbolic Music

Marina Borsodi-Benson (EPFL); Martin Rohrmeier (EPFL)

Galant-style voice leading schemata are specific musical patterns that describe two- to four-voice structures, characterized by their voice-leading, harmonic, rhythmic, and metric profiles. While schema descriptions are straightforward on their own, in real musical contexts they are highly elaborated. Consequently, automatic labeling is not only a matter of surface pattern matching within a polyphonic context, but also identifying the underlying structure of an elaborated sequence. This paper formulates schema identification as a subgraph matching problem. A musical piece is modeled as a directed graph where nodes represent notes and edges represent the musical relationships between notes. These edges encode both horizontal (voice-leading) and vertical (harmonic) intervals, weighted by temporal proximity. Schemata are similarly encoded as graph templates. Using this graph architecture, a subgraph matching algorithm identifies schema candidates within the piece graph, which are subsequently filtered by edge weight and metric balance. To evaluate this approach, we test the algorithm on a corpus of Mozart Piano Sonatas with expert schema annotations. Preliminary results indicate that this graph-theoretic model identifies many schemata successfully, despite surface ornamentation. We also successfully limit the number of false positives compared to previous approaches. This framework demonstrates how graph-theoretic models can be generalized to identify complex, highly elaborated musical patterns in polyphonic symbolic music.

Analyzing Key Changes in Three Corpora of Popular Music

Jenine Brown (Peabody Conservatory of Music)

When music changes key at the end of a pop song, these modulations (i.e., key changes) are routinely described by cognition scholars (Huron, 2006; Levitin, 2024) and music theorists (Hanenberg, 2016; Everett, 1997; Doll, 2011; Griffiths, 2015; Osborn, 2013; Ricci, 2017) as imbuing energy into the song. The frequency and intervallic distance of these key changes, however, has not been studied systematically. In response, this project analyzed three corpora of rock/pop music to determine 1) what percentage of songs contain modulations, 2) how many modulations occur per song, and 3) which modulation intervals are most common. To facilitate this study, the three popular music corpora were analyzed: the Rolling Stone Corpus (de Clercq & Temperley, 2011) and two that I created: the Pentatonix corpus tracks modulations in all songs by the a cappella group, and the David Foster corpus does the same for all songs in which the composer/arranger acted as producer/writer. These varied corpora allow for comparison as to whether modulation frequency/type differs among them. Results are that in all three datasets, modulating just once happens more than modulating many times in a single song and the most frequent modulation is up by m2, with up by M2 as the second-most frequent. These results also suggest that the Pentatonix modulate more often, suggesting a compositional hallmark of the a cappella group. Other analyses will be shared, including the most frequent formal location for modulations and whether songs with modulations are disproportionately likely to have won Grammy awards. This research is the first to offer data on modulations in popular music. Findings are of interest to many music-academic fields, ranging from enhancing understandings of modulations in music theory, to providing ideas for future music-cognitive research, to informing music composition/arranging, and to developing more robust corpus-based music analyses of rock/pop music.

References

[1] de Clercq, T., & Temperley, D. (2011). A corpus analysis of rock harmony. Popular Music, 30(1), 47–70.
[2] Doll, C. (2011). Rockin' out: Expressive modulation in verse-chorus form. Music Theory Online, 17(3).
[3] Everett, W. (1997). Swallowed by a song: Paul Simon's crisis of chromaticism. In J. Covach & G. M. Boone (Eds.), Understanding rock: Essays in musical analysis (pp. 113–153). Oxford University Press.
[4] Griffiths, D. (2015). Elevating form and elevating modulation. Popular Music, 34(1), 22–44.
[5] Hanenberg, S. J. (2016). Rock modulation and narrative. Music Theory Online, 22(2).
[6] Huron, D. (2006). Sweet anticipation: Music and the psychology of expectation. MIT Press.
[7] Levitin, D. J. (2024). I heard there was a secret chord: Music as medicine. W. W. Norton & Company.
[8] Osborn, B. (2013). Subverting the verse/chorus paradigm: Terminally climactic forms in recent rock music. Music Theory Spectrum, 35(1), 23–47.
[9] Ricci, A. (2017). The pump-up in pop music of the 1970s and 1980s. Music Analysis, 36, 94–115. https://doi.org/10.1111/musa.12086

CAMAT_v2: Cloud-Based Tools for Symbolic Music Analysis

Egor Polyakov (HfM Weimar); Martin Pfleiderer (HfM Weimar)

Computational musicology has increasingly moved into web-based environments. Tools such as Verovio and mei-friend have made digital editions accessible in the browser, while analytical workflows for symbolic music often still depend on local installations, heterogeneous file formats, and project-specific scripts. This presentation introduces the public release of an open-source Python toolkit for sheet-music analysis, developed in the DFG project "A Comprehensive Cloud-Based Toolbox for Sheet Music Analysis" at the University of Music Franz Liszt Weimar. The toolkit is designed for research in musicology and music theory with common music notation, while also supporting experimental repertories such as spectral/graphic notation, drum notation, rap transcriptions, and mensural notation. It is divided into two packages. The first package, camat_corpus, provides comprehensive tools for preparing research corpora from OMR output and digital editions, with workflows for MEI editing, IIIF facsimile handling and synchronization, validation, and the compilation of reusable datasets. The second package, camat_v2, provides a unified analytical layer over several parsing backends, including Verovio, music21, and Partitura, enabling conversion and processing of MEI, kern, MusicXML, and related formats. It also includes a sophisticated MEI-based annotation framework that connects Python-based analysis with MEI and mei-friend workflows. Beyond import and rendering, camat_v2 offers a statistical stack for analyzing pitch and rhythmic distributions as well as motif- and texture-search routines based on binary, piano-roll-like representations of musical structures. This approach connects established pitch-time and multidimensional representations for pattern discovery with recent MIR and machine-learning work on matrix-based symbolic inputs. The presentation will introduce these toolkits and demonstrate example analyses in Jupyter and mei-friend using corpora assembled within the project. In particular, we will show Jupyter4NFDI, a dedicated cloud-based notebook service coordinated by Forschungszentrum Jülich/JSC, which provides browser-based access to Python- and Jupyter-based workflows without requiring local software installation.

References

[1] Cancino-Chacón, C. E., Peter, S. D., Karystinaios, E., Foscarin, F., Grachten, M., & Widmer, G. (2022). Partitura: A Python package for symbolic music processing. In Proceedings of the Music Encoding Conference (MEC 2022).

[2] Cuthbert, M. S., & Ariza, C. (2010). music21: A toolkit for computer-aided musicology and symbolic music data. In Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010) (pp. 637-642).

[3] Goebl, W., & Weigl, D. M. (2024). mei-friend: An interactive web-based editor for digital music encodings. Journal of Open Source Software, 9(98), 6002. https://doi.org/10.21105/joss.06002

[4] Jülich Supercomputing Centre. (n.d.). Jupyter4NFDI. Forschungszentrum Jülich. Retrieved May 30, 2026, from https://www.fz-juelich.de/en/jsc/projects/jupyter4nfdi

[5] Müller, M., & Zalkow, F. (2019). FMP notebooks: Educational material for teaching and learning fundamentals of music processing. In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR 2019) (pp. 573-580).

[6] Pugin, L., Zitellini, R., & Roland, P. (2014). Verovio: A library for engraving MEI music

notation into SVG. In Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014) (pp. 107-112).
[7] Zhang, H., Karystinaios, E., Dixon, S., Widmer, G., & Cancino-Chacón, C. E. (2023). Symbolic music representations for classification tasks: A systematic evaluation. In Proceedings of the 24th International Society for Music Information Retrieval Conference (ISMIR 2023) (pp. 848-858). https://doi.org/10.5281/zenodo.10265421

Assessing Rhythm Complexity Measures Using the Item Response Theory

Annalena Rusch (LORIA); Jonas Bertoli (LORIA); Florence Levé (MIS); Geoffray Bonnin (LORIA); Anicet Bart (Solfy)

Rhythm production is among the key skills to acquire in music theory. As any skill, its learning is facilitated by progressively confronting students with rhythm sequences of increasing complexity. Although several complexity measures have been proposed in the literature (Povels and Essens, 1985; Pressing, 1999; Thul and Toussaint, 2008), their relevance is still unclear. To address this issue, psychometric frameworks can be leveraged to ground these metrics in student performance. One prominent approach in this context is the Item Response Theory (IRT) (Hambleton and Swaminathan, 1984), which models the probability of successfully completing an activity as a function of students' latent characteristics and activities' latent characteristics. Applied to rhythm production, the activities latent characteristics largely correspond to the complexity of the underlying sequences, but also include other aspects such as the proximity with the students' culture. We therefore propose to rely on the correlation with these latent characteristics as a means of assessing the relevance of a complexity measure. Since IRT models rely on performance data as input, we also propose a new measure of performance on rhythm production tasks. We collected 5 months of data from 716 students who completed rhythm production activities on 4909 different sequences using the music learning platform Solfy. After filtering out outlier students and keeping only activities completed more than 50 times, 624 students and 91 sequences remained. We experimented with different complexity measures from the literature using the resulting dataset. Our results show that the most relevant measure is the measure of Pressing (Pressing, 1999), which led to a statistically significant Pearson correlation of 0.4 with the sequences' IRT characteristics. A qualitative analysis of the sequences with extreme characteristics also revealed the extent to which the distinction between the precision of the physical execution and the understanding of the rhythm patterns matters.

Is the Musical Mind Full of Trees? Probing the Validity of Hierarchical Harmonic Sensitivity as an Indicator of Musical Ability

Evan Hertafeld (University of Amsterdam); John Ashley Burgoyne (University of Amsterdam); Xiaoxuan Wang (EPFL)

We implement a musical context-free grammar to hierarchically parse a corpus of classical music at the level of harmonic annotations. Production rules encode musical concepts such as phrasal structure, functional theory, and modulation. Computational analysis of the resulting structures identifies musical sections which have either a strong or weak hierarchical dependence on their harmonic context. We combine this with an ngram-based model that identifies sections where local expectation is violated. By combining these models, we can select musical areas which are all equally surprising with a strictly local context, but vary in their hierarchical depth. From these we generate a musical questionnaire to identify whether musical ability correlates with chord identification of hierarchically deep sections of music. In total, 11.074 musical phrases were parsed from a corpus of 1.283 classical movements, resulting in a dataset of 822 million parse trees. Depending on particular selection criteria, our approach identifies in between 500 and 1.000 musical sections containing a pivotal chord which is locally surprising and of stable hierarchical depth, and thus suitable for inclusion in a musical questionnaire. Vitally, our approach does not depend on hand-annotated treebanks, but rather operates on the entire parse chart by selecting chords whose depth is statistically stable across valid parses. Our analysis thus explores the degree to which the aggregate tendencies of the structures enforced by musical functions have representational correlates in a musical listener. If this correlation is found, our process offers a robust means to quickly generate assessments of hierarchical musical ability across a wide range of musical genres and styles without the need for expensive hand-selection by experts. Such assessments could prove useful tools for a wide range of work in musical research.

References

[1] Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. Journal of Mathematics and Mu- sic, 5(1), 35–53. Pearce, Marcus Thomas. The construction and evaluation of statistical models of melodic structure in music perception and composition. Diss. City University London, 2005.

A Music-based AR Game for Upper-Limb Motor Rehabilitation: The Effect of Rhythmic Cues on Engagement and Performance

Sanne Duinker (Utrecht University); Hanna J. Hauptmann (Utrecht University); Anja Volk (Utrecht University)

Recent years have seen a growing interest in the use of music-based approaches in rehabilitation, particularly for motor rehabilitation [1]. One of the key challenges in rehabilitation concerns individuals experiencing a lack of motivation for repeated at-home exercises. To increase motivation, studies have explored the use of augmented reality (AR), enhancing engagement through gamification and immersion. At the same time, music-based technologies have been developed to improve engagement through rhythmic cues employing the principle of entrainment [1-4]. However, there is a gap in combining AR with rhythmic cues, specifically for upper-limb motor rehabilitation. We developed and evaluated a music-based AR game to assess the effect of rhythmic cues on user engagement and task performance for upper-limb motor rehabilitation. To investigate the effects of rhythmic cues in the AR game, musical pieces with a clear beat structure and different tempi were employed, building on previous clinical studies investigating the role of tempo [5]. We employed three different tempo conditions and a no-music condition, to investigate if the presence of rhythmic cues has an effect on performance and engagement, and whether the effect varies across different tempi. A within-subject study was conducted with twenty healthy participants, gathering gameplay data for assessing task performance, and user engagement data via a questionnaire accompanied by a post-interview. The task performance metrics showed that participants used the tempo to guide their movements, with 100 BPM being the most 'natural' tempo. Post-interviews revealed that engagement was influenced by mastery and enjoyment, with the music conditions being experienced as more enjoyable than the no-music baseline. This study shows that rhythmic cues in an AR upper-limb motor rehabilitation game can enhance enjoyment and guide movement. Future research could tailor tempo or song to the user's preference, and investigate musical structures beyond rhythm cues to achieve optimal personalized engagement for rehabilitation.

References

[1] P. R. Kantan, S. Dahl, H. R. Jørgensen and E. G. Spaich, "Making Movement Sonification Usable in Clinical Gait Rehabilitation: A User-Centered Study," in AM '24: Proceedings of the 19th International Audio Mostly Conference: Explorations in Sonic Cultures, Milan, Italy, 2024.

[2] T. Y. Pang, T. Connelly, F. Feltham, C.-T. Cheng, A. Rahman, J. Chan, L. McCarney and K. Neville, "A Wearable Personalised Sonification and Biofeedback Device to Enhance Movement Awareness," MDPI, 2024.

[3] B. Luciani, A. Pedrocchi, P. Tropea, A. Seregni, F. Braghin and M. Gandolla, "Augmented reality for upper limb rehabilitation: real-time kinematic feedback with HoloLens 2," Springer Nature Link, 2025.

[4] J. A. Grahn, "Neural Mechanisms of Rhythm Perception: Current Findings and Future Perspectives," TOPICS, vol. 4, no. 4, pp. 585-606, 18 July 2012.

[5] M. Leman, D. Moelants, M. Varewyck, F. Styns, L. van Noorden and J.-P. Martens, "Activating and Relaxing Music Entrains the Speed of Beat Synchronized Walking," PLOS One, 2013.

A Data Set for Chant Generation and Analysis

Geert Maessen (Gregoriana Amsterdam)

This poster presents data published on GitHub. Below a general description that will be specified on the poster. Medieval Europe had various liturgical chant traditions with interconnections. Best known is Gregorian chant. The longest chants (offertoria or sacrificia) from various traditions have been collected in the data. Besides Gregorian (GRE), these include Old Roman (ROM), Milanese (MIL), and Beneventan (BEN). From the sixth untill the eleventh century the Mozarabic rite existed on the Iberian peninsula. When pitch readable notation came in use (in the eleventh century) this rite was replaced by the Roman rite with its Gregorian chant. Thousands of chants have been preserved in neumatic notation in which we cannot read intervals. Its most important manuscript is the tenth-century León antiphonary. LEO is the collection of all León sacrificia (offertories). CIS is the collection of sacrificia from the sixteenth-century record of the orally transmitted Mozarabic rite: the Cisneros Cantorales. HIS is the collection of the few known pitched melodies before 1200. GRE, ROM, MIL, BEN, CIS, and HIS are encoded in Volpiano. LEO is encoded in contour notation. Both will be explained on the poster. Two more sets are included. AMS-1: all Gregorian chants that exist with the same text in León, and its invers AMS-2: all chants from León with the same text as Gregorian chants. AMS is: Antiphonale Missarum Sextuplex (1935), Hesbert's edition of all Gregorian chants from the six oldest (ninth-century) mass manuscripts, compared with Randel's Index of Mozarabic chants (1973). With this data, we hope to improve research into the relationship between medieval chant traditions and encourage further research into melody generation for the Mozarabic rite. We think that apart from Maessen's alignment method (2017), Conklin's statistical method (2019), and Camas' proposed phylogenetic method (2024), there may be other methods as well.

References

[1] René-Jean Hesbert (1935), Antiphonale Missarum Sextuplex, Herder, Rome.
[2] Don Randel (1973), An Index to the Chant of the Mozarabic Rite. Princeton University Press.
[3] Geert Maessen & Peter van Kranenburg (2017), A Semi-Automatic Method to Produce Singable Melodies for the Lost Chant of the Mozarabic Rite. In Proceedings of the 7th International Workshop on Folk Music Analysis, Málaga, Spain.
[4] Darrell Conklin & Geert Maessen (2019), Generation of melodies for the lost chant of the Mozarabic rite, Applied Sciences, 9(20).
[5] Francisco M. Camas, Tania Toledano & Paloma Gutiérrez del Arroyo (2024), Phylogeny and melodic recovery of modal melismas in early Franco-Roman antiphoner of Silos, International Conference on the Intersection of Digital Technologies and Music Research, Lisboa, Portugal.

A Corpus Study of Form and Texture in Cantopop (1975–1999): Evaluating the Influence of Cover Songs

Jason Yin Hei Lee (McGill University/CIRMMT)

While the historical development of Cantonese popular music (Cantopop) has been well documented (Chu 2017; Wong 2014), the evolution of its musical style remains insufficiently understood due to a lack of systematic studies. To bridge this gap, this paper presents a corpus study that tracks Cantopop's stylistic evolution around the 1980s, commonly regarded as the golden age of the genre and marked by an influx of globally sourced cover songs (Chu and Leung 2013). By providing empirical data on form and texture, this study demonstrates how this influx transformed the genre's musical style. Expanding the corpora developed in my previous work (Lee 2023, 2026), this paper broadens its historical scope to include songs released from 1975 to 1999. Drawing from London's (2013) methodology for constructing a balanced, multi-source corpus, I selected 10 songs per year based on various factors, such as the songs' commercial success and critical reception. Formal sections and textural features of the songs were encoded following Reymore et al.'s (2022, 2025) methodology. By contrasting cover and original tracks in the corpus, I analyze the formal and textural encodings to report on their respective trends. First, the influx of cover songs correlates with a general shift in song form, as verse-chorus form replaces AABA as the predominant song form during the 1980s. By comparing my data with corpus studies of Anglo-American pop-rock (Ensign 2015; Summach 2012), my study substantiates claims of Cantopop's stylistic stagnation in the 1970s (e.g., Ogawa 2001) and suggests that cover songs constituted a key strategy for a stylistic breakthrough. Second, the corpus data show that the influx of cover songs coincided with shifts in textural and instrumentation norms. By tracking the deployment of the "explicit beat layer" in the texture (Moore 2012), I demonstrate how cover songs introduced new instrumentation practices that contributed to sub-genre differentiation within Cantopop.

References

[1] Chu, Yiu-Wai. 2017. Hong Kong Cantopop: A Concise History. Hong Kong University Press.

[2] Chu, Yiu-Wai, and Eve Leung. 2013. "Remapping Hong Kong Popular Music: Covers, Localisation and the Waning Hybridity of Cantopop." Popular Music 32 (1): 65–78.

[3] Ensign, Jeffrey S. 2015. "Form in Popular Song, 1990–2009." PhD diss., University of North Texas.

[4] Lee, Jason Yin Hei. 2023. "A Corpus Study of Tone-Melody Correspondence in Cantopop, 2000–2020." MA thesis, University of British Columbia. https://doi.org/10.14288/1.0431335.

[5] Lee, Jason Yin Hei. 2026. "Stylistic Evolution of Cantopop: Formal Trends in 1980–2024." Conference paper presented at Hong Kong as Music Theory Conference, University of Hong Kong, June 12.

[6] London, Justin. 2013. "Building a Representative Corpus of Classical Music." Music Perception 31 (1): 68–90. https://doi.org/10.1525/MP.2013.31.1.68.

[7] Moore, Allan F. 2012. Song Means: Analysing and Interpreting Recorded Popular Song. Ashgate.

[8] Ogawa, Masashi. 2001. "Japanese Popular Music in Hong Kong: Analysis of Global/Local Cultural Relations." In Globalizing Japan Ethnography of the Japanese Presence in Asia, Europe, and America, edited by Harumi Befu and Sylvie Guichard-Anguis. Routledge.

[9] Reymore, Lindsey, Nicole Biamonte, Ben Duinker, et al. 2025. “Analyzing Timbre and Texture in a Multi-Genre Popular Music Corpus.” Conference paper presented at the Third International Conference on Computational and Cognitive Musicology, Aalborg, Demark, October 10.
[10] Reymore, Lindsey, Ben Duinker, Matthew Zeller, et al. 2022. “Encoding and Analyzing the Timbre in Popular Song (TiPS) Corpus.” In Proceedings of the Music Encoding Conference 2022, edited by Ai Lynn Ang, Jennifer Bain, and David M. Weigl.
[11] Summach, Jay. 2012. “Form in Top-20 Rock Music, 1955–1989.” PhD diss., Yale University.
[12] Wong, Chi Wah 黃志華. 2014. 原創先鋒—粵曲人的流行曲調創作 Jyuncong Sinfung—Jyutkukjan Dik Lauhangkukdiu Congzok [Pioneers of Originality: Popular Tunes from Practitioners of Cantonese Opera]. Joint Publishing.

Identifying Structural Principles Across Musical Systems Through Generative Modeling

Christoph Finkensiep (EPFL)

Music research across different styles is faced with a dilemma: On one hand, music is made by humans with roughly the same biological hardware. On the other hand, musical styles and traditions across time and cultures exhibit an astonishing variety that hardly exhibits any common patterns. This raises the question of what constitutes the link between human cognition and the variety of actual musical expressions. Current research on this mostly focuses on basic features of music(s) that are so general that they can reasonably be ascribed to most musics in the world, as well as statistical universals (Jacoby et al. 2020). However, this approach offers little for explaining more complex, higher-level structures that are specific to certain styles. The goal of this talk is to propose an alternative approach that identifies structural principles through generative modeling. This approach will be discussed on the example of melodies. Melodies are a popular object of study because they are very wide-spread and do not require many specific assumptions, and because they exhibit a simple surface structure: only one pitch sounds at a time and notes are ordered in a sequence with direct predecessors and successors. However, the underlying structure of melodies can be much more complex: notes can be related to other notes in a non-local fashion, for example through a hierarchical organization, implicit polyphony, latent harmonic structure, or combinations of these. Generative models explain the structure of musical artifacts such as melodies in terms of underlying generative principles which may be shared across styles but used and combined in different ways. This approach is complementary to universals: It starts with a particular style but can still reveal common principles across styles. These principles are cognitive tools that are generally available but (in contrast to universals) optional.

References

[1] Jacoby, Nori, Elizabeth Hellmuth Margulis, Martin Clayton, et al. 2020. "Cross-Cultural Work in Music Cognition: Challenges, Insights, and Recommendations." Music Perception 37 (3): 185–95. https://doi.org/10.1525/mp.2020.37.3.185.

Between Graph Semantics and Musicological Intent: Ambiguity in Researcher–Machine Communication Over Music Knowledge Graphs

Polina Proutskova (Industry Commons Foundation); Olga Velichkina (French Society for Ethnomusicology); Harshdeep Singh (Odoma); Matteo Romanello (Odoma)

Music knowledge graphs promise new forms of exploration across collections, repertoires, recordings and metadata. Yet interaction between a musicologist and a graph-based system is not simply translation from natural language into SPARQL. Researchers ask conceptual and interpretive questions, while datasets encode knowledge through formal classes, relations, provenance and modelling decisions. This gap creates ambiguity: does "influence" mean citation, stylistic similarity, teacher–student transmission, shared repertoire or documented contact? Does "where a musical encounter took place" mean venue, city, textual mention or inferred event location? This talk proposes a framework for studying ambiguity in researcher–machine communication over music knowledge graphs. We approach the problem from two directions. First, in the Musparql pipeline, existing human-authored SPARQL queries serve as graph-grounded starting points from which natural-language questions are generated, aligned with evidence from papers, documentation and sources, then reviewed, rewritten and annotated. Second, in experiments with the GRAPHIA agent for federated knowledge graphs, including LinkedMusic, Culture Knowledge Graph, Polifonia knowledge graphs and Wikidata, musicologists pose questions that cross dataset boundaries, for example placing musical works in relation to local contexts, documented encounters and composers' biographies; the agent's responses and user journeys are evaluated. The framework treats ambiguity as evidence of where formal music-data structures and musicological concepts fail to align. As in music transcription, disagreement and alternative formulations may reveal interpretive assumptions rather than error. Early examples show tensions between formal completeness and musicological salience: provenance fields may be returned but communicatively secondary; aggregation values may be operationally necessary but not central. Crucially, phrasing adequacy often depends on ontology and dataset knowledge: the same returned field may be routine metadata in one context and decisive evidence in another. The contribution is a methodological and epistemological framework for studying how musicological meaning is negotiated between datasets, formal representations, LLM agents and human researchers.

(In)Calculable Humanness in Artificial Music-Making: Unhyping AI in Computational Ethnomusicology After the Digital

Yinuo Chen (King's College London)

It may be wearying to invoke AI (again), yet this position paper does so knowingly: to unhype it is already to hype it. It asks what computational ethnomusicology prefigures when musical life is not merely mediated by platforms but formatted through computation, and when, under generative AI, the human is in, out of, or rendered spectral within the loop. This paper begins from a paradox: generative systems promise to save creative time, yet they generate new burdens of justification. The producer is not freed from labour but asked, implicitly and explicitly, to prove that the work holds skill, intention, taste, and humanness. This poses an epistemological problem for ethnomusicology as computation enters musical practice and research. When musical life is digitised, platformised, and at once bifurcated by and sympoietic with generative systems, authorship becomes a question posed twice over: who, or what, is recognised as the creator of the musical object, and who, or what, does the ethnomusicologist, with the same tools, constitute as the creating subject of study? Empirically, the paper undertakes a computationally attentive digital ethnography of AI music discourse in China and its creator economy, pairing FCDA (feminist critical discourse analysis) with engagement-metric analysis across viral videos, tracing how producers use, explain, promote, and contest tools like Suno and Udio while accentuating/diminishing the human within the artificial, finding legitimacy accrues less to craft than to visible explanatory labour. Drawing on platform studies scholarship, feminist STS, and posthuman theory, the paper presses computational ethnomusicology beyond sound, style, or corpus toward the apparatuses that condition and configure visibility, and treats authorship as an effect enacted across assemblages of human and nonhuman actants, where the creator of the musical object is decided rather than presupposed, and asks if humanness itself is rendered calculable through the agential cut of measurement.

Comparing Harmony Across Diverse Datasets: Characteristic Cadences in Brazilian Choro, Jazz, Popular and Western Classical Music

Mauro Orsini Windholz (University of Würzburg); David R. W. Sears (University of Michigan); Fabian C. Moss (University of Würzburg)

While corpus studies on musical harmony have increased our knowledge on the harmonic behavior of different styles based on large databases, they rarely perform direct, data-driven comparisons between styles. Moreover, such studies rarely include styles from the Global South. This presentation discusses results of studies comparing triadic harmonic progressions at moments of formal segmentations across databases of harmonic annotations in Brazilian choro, Anglophone popular, European classical and jazz music. The methodology drew on and expanded the methods in Sears and Forrest, 2021. Tri-grams of Roman Numerals ending on strong formal boundaries, according to metrics derived from IDyOM (Pearce, 2005), were filtered from each dataset. From this filtered list, two metrics were calculated for each tri-gram: scaled pointwise mutual information, which measures the conventionality of tri-grams within a style, and relative frequency ratio, which measures how characteristic a tri-gram is of a style when directly compared with another style. Using the characteristicness values obtained, comparisons between three styles at once were carried out: choro, popular and classical and choro, popular and jazz. Results suggest that Brazilian choro harmony at formal boundaries features characteristically higher use of applied dominants than all three styles, and that certain harmonic sequences are unique to choro when compared to the other styles, such as iv - iiø - i. While jazz also features high use of applied dominants, these are overwhelmingly nested within applied 2-5-1 patterns, a different use of applied dominants than in choro. These results show that computationally-driven comparisons of different styles can help music theory determine, based on relatively large amounts of data, what are the unique harmonic characteristics of different styles and how they compare with each other. Moreover, investigating styles from diverse contexts helps broaden the scope of harmonic devices under study.

Poster Session B

Discussing Representation Challenges for Orchestral AMT Using the Beethoven Symphonies Excerpt Dataset

Hans-Ulrich Berendes (International Audio Laboratories Erlangen); Abhirup Saha (International Audio Laboratories Erlangen); Ben Maman (International Audio Laboratories Erlangen); Vlora Arifi-Müller (International Audio Laboratories Erlangen); Meinard Müller (International Audio Laboratories Erlangen)

Automatic music transcription (AMT) has increasingly moved from controlled settings such as solo piano toward highly polyphonic orchestral music. This shift exposes fundamental problems concerning the representation of music across symbolic scores and audio recordings. In this contribution, we use the Beethoven Symphonies Excerpt Dataset (BSED), a cross-version dataset containing aligned score representations and multiple performances of excerpts from Beethoven's symphonies, as a case study for discussing these issues. Our contribution has a twofold focus. First, we discuss why the curation of orchestral music datasets is inherently difficult and why aligning scores, note-event representations, audio recordings, and multiple performances remains challenging. Existing formats such as MusicXML, MIDI, and note-event lists provide convenient abstractions, but only partially capture the complexity of orchestral music. Dense polyphony, onset asynchrony, performance differences between versions, and unison passages challenge basic assumptions about simultaneity, note identity, and voice separation. Using BSED as a case study, we take these challenges as a starting point for discussing computational representations of orchestral music. Second, we discuss AMT as one of the main applications motivating the design of BSED while simultaneously questioning current transcription paradigms in orchestral settings. Standard AMT evaluation assumes a well-defined correspondence between annotated score events and transcription targets, an assumption that often breaks down in orchestral music. For example, different strategies for handling unison notes and onset asynchronies can lead to substantially different evaluation results for identical transcription outputs. We therefore argue that orchestral AMT is not only technically challenging, but partially ill-defined as a task itself, motivating broader discussions about computational representations of music.

Human vs AI in Symbolic Music: An Online Listening Study on Bach-Style Chorales

Jason Ackermann (University of Würzburg); Ricarda Ernsperger (University of Würzburg); Fabian C. Moss (University of Würzburg)

The field of AI music generation is developing rapidly, and models have recently achieved notable results [1, 2]. These trends raise the question to which degree human and AI compositions can be distinguished in everyday scenarios. We present an online listening study evaluating the human ability to differentiate AI-generated chorales from human-composed works across various listener demographics. We generate chorales using NotaGen [1] and compare them with chorales by J. S. Bach taken from the music21 corpus [3]. Participants listened to isolated musical phrases and were asked to categorize the snippets as AI or human composed. Furthermore, we collect additional data such as age, musical experience or AI usage in order to contextualize the results. Preliminary results (N = 17) suggest substantial variability in discrimination accuracy across participants. Prior musical experience does not appear to improve classification performance, while participants with less frequent AI usage tend to achieve higher discrimination accuracy. Ongoing work aims to validate these findings with a larger participant cohort and a wider range of musical excerpts.

References

[1] Yashan Wang et al. NotaGen: Advancing Musicality in Symbolic Music Generation with Large Language Model Training Paradigms. 2025. arXiv: 2502.18008 [cs.SD]. url: https://arxiv.org/abs/2502.18008.

[2] Xingwei Qu et al. MuPT: A Generative Symbolic Music Pretrained Transformer. 2024. arXiv: 2404.06393 [cs.SD]. url: https://arxiv.org/abs/2404.06393.

[3] Music21.Corpus.Chorales — Music21 Documentation. url: https://music21.org/music21docs/moduleReference/moduleCorpusChorales.html (visited on 05/19/2026)

Conceptualizing Structural Repetition in Tonal Music

Zeng Ren (EPFL); Xinyi Guan (EPFL); Martin Rohrmeier (EPFL)

Understanding how music repeats is a central challenge in music theory, music cognition and music information retrieval (MIR). Musically interesting repetitions occur in varied forms, across levels of reduction, and govern long timespans and different musical parameters. Existing notions of musical repeats, together with their corresponding processing models can only account for a small subset of the aforementioned phenomena. How could we conceptualize musical repeats to account for all of them? What computations are involved in processing such repetition? This article presents a structural account of musical repeats in terms of how patterns are represented, abstracted, and instantiated in varied forms. A central claim is that musical patterns are composable elaboration procedures on musical events. Through an extensive reductive analysis on the third movement of Beethoven's Piano Sonata Op. 57 ("Appassionata"), we demonstrate that 1) elaboration procedures can capture different hearings and usages of the same musical surface, 2) a single elaboration procedure can re-occur across levels of reduction, and 3) variations can be decomposed into smaller reusable patterns. More broadly, the article provides new perspectives on musical repetition for different disciplines. For MIR, the framework offers a theory-informed formulation for pattern discovery beyond the musical surface. For music cognition, it hypothesizes a representation of musical patterns that can support categorization and generalization. For music theory, it provides a formal description of how repetition structure interacts with a hierarchical tonal structure, which explains not only what is repeated, but also how the repeated entities are put together coherently.

Compression-Based Modeling of Similarity in Jazz Harmony

Xinyi Guan (EPFL); Edward Hall (EPFL); Zeng Ren (EPFL); Marina Borsodi-Benson (EPFL); Claire Pelofi (NYU); Martin Rohrmeier (EPFL)

The ability to accurately account for musical similarity is essential to a wide range of tasks in MIR and cognition. Compression distance provides a promising and underexplored model for similarity: two pieces are similar to the extent that a compact description of one helps describe the other. This paper investigates two competing compression strategies, focusing on surface and structural regularities, respectively. The first estimates how predictable one sequence is under a probabilistic model trained on another using the Prediction by Partial Matching algorithm (Cleary & Witten, 2003). The second approach measures how much a piece's derivation tree can be simplified with the learned abstractions from another, using a library-learning algorithm, Stitch (Bowers et al., 2023). Given a collection of derivation trees, it treats each instance as a program and compresses the collection by learning a library of reusable program abstractions that capture recurring substructures and rewriting the programs in terms of these abstractions. We apply both models to a corpus of jazz harmony analyses and measure their compression performance, comparing interaction between the two. We further present qualitative analyses to demonstrate where the diverge and agree. The comparison clarifies what each model treats as relevant regularity: the PPM is more sensitive to superficial transitions, while Stitch operates at a higher representational level, discovering reusable patterns and showing greater robustness by not considering all transitions equal.

Evaluating Mozart's K. 331: A Computational Stylometric Approach to the Chronological Controversy

Zhenyu Jiang (The Chinese University of Hong Kong, Shenzhen)

Wolfgang Amadeus Mozart's celebrated Piano Sonata No. 11, K. 331 is familiar to almost every piano learner. Its striking use of variation form in the first movement and adoption of Turkish elements in the famous rondo finale makes it an essential work in piano literature. However, due to the scarcity of contemporary documentation, the compositional date of this sonata remains a chronological debate in musicological discourse. While Köchel suggested 1778 by documentary evidence such as Mozart's correspondence, other scholars, such as Tyson, Eisen, Sadie, and Plath etc., proposed different chronology by source criticism or stylistic analysis, narrowing it to 1781-1784. Besides, Irving (1997) suggested that Mozart used soprano clef in these sonatas, which was unusual in Mozart's works and usually used in counterpoint textbook. Thus, these sonatas were likely pedagogically intended. By translating qualitative stylistic observations into computable musical features, this study explores how quantitative analysis may contribute to these existing debates concerning the chronology of K. 331. Mijangos et al. (2024) successfully quantified stylistic differences in string quartets by Haydn, Mozart, and Beethoven and found results are consistent with the standard view of Haydn as the initiator of the genre and Beethoven as the most innovative among the three, using sustained homology. Inspired by that, this study will employ a comparative corpus-based approach. Two groups with approximately 8-10 works per period of Mozart's keyboard works using Köchel's chronology, specifically piano sonatas and variation sets, will be the corpus: Group A (1777–1778, Paris period) and Group B (1782–1784, early Vienna period). Utilizing the music21 Python library and KernScores database, high-dimensional musical features, including harmonic N-grams, rhythmic density, and melodic interval distributions will be extracted. By analyzing it in a computational way, this study aims to explore how quantitative methods contribute to traditional analysis and connect to historical reality.

Extracting Beats and Downbeats from Classical Music Audio: Investigating State-of-the-Art Methods

Lele Liu (University of Würzburg); Christof Weiß (University of Würzburg)

Beat and downbeat tracking are fundamental tasks in the field of music information retrieval. Recent research has shown that beat tracking methods achieve strong results on popular music with relatively stable tempo, but substantially worse results on classical music, where expressive performances introduce significant tempo variation [1, 2]. Furthermore, downbeat tracking remains challenging regardless of the music genre. In this work, we focus on the challenging task of downbeat tracking in classical music. We evaluate state-of-the-art beat and downbeat tracking models on multiple classical music corpora to analyse the limitations of current methods. We evaluate on datasets including the Schubert Winterreise Dataset, the Beethoven String Quartets Dataset, the RWC Classical Dataset, and the Wagner Ring Dataset. We also evaluate on an unseen dataset (Rubato) that broadly captures classical music in various performance versions. Some of these datasets contain only downbeat annotations without beat annotations, which makes the task even more challenging. We evaluate several open-source beat and downbeat tracking models, including temporal convolutional networks [3], a Transformer-based method [1], and the cross-modal beat tracking model proposed in [2]. The models are evaluated across multiple classical music corpora under both dataset-specific and cross-dataset generalisation settings. Our results show that current models still struggle to capture beats and downbeats in classical music audio. In particular, downbeat tracking in classical music remains extremely challenging across all evaluated datasets. We identify several scenarios in which beat and downbeat tracking models commonly fail, including performances with extremely slow tempi, string instrument textures, and significant rubato. These analyses highlight the existing limitations of existing beat and downbeat tracking methods and bring useful insights for improving beat and downbeat tracking systems for classical music.

References

[1] F. Foscarin, J. Schlüter, and G. Widmer, "Beat this! accurate beat tracking without dbn postprocessing," in International Society for Music Information Retrieval Conference, 2024.
[2] C. Chiu, L. Liu, C. Weiß, and M. Müller, "Cross-modal approaches to beat tracking: A case study on chopin mazurkas," accepted to Transactions of the International Society for Music Information Retrieval, 2025.
[3] S. Böck and M. E. P. Davies, "Deconstruct, analyse, reconstruct: How to improve tempo, beat, and downbeat estimation," in Proceedings of the International Society for Music Information Retrieval Conference, (Montreal, Canada), pp. 574–582, 2020.

Tonal Organization in Monophonic Jazz Solos: Putting Chord Scale Theory to the Test

Klaus Frieler (Max Planck Institute for Empirical Aesthetics, Frankfurt/Main); Tim Eipert (University of Würzburg); Fabian Moss (University of Würzburg)

Chord-scale theory (CST) is a widely used framework in jazz theory and pedagogy for associating chords with scales. Despite its prevalence, CST has not been empirically evaluated in the context of jazz history. The precise relationship between chords and scales remains elusive because of the tension between vertical (chordal) and horizontal (melodic) structures, as well as the influence of local and global tonal contexts on pitch selection. This study investigates the tonal organization of 444 tonal, non-modal jazz solos from the Weimar Jazz Database by analyzing pitch choices in their relation to underlying chords and scale degrees. Using algorithmic matching techniques, we identify the most common scales associated with all chord types across the corpus. Additionally, we present three case studies to explore matters in greater detail: (1) the treatment of tritone substitutes for dominant chords, (2) the scalar interpretations of fully diminished chords, and (3) the use of blue notes. Our findings show that tonal choices in historical jazz solos are grounded in the chord tones ("laying out the chord"). These are enriched, first, with diatonic pitches, mostly chosen according to the tonal context given by the local scale degree of the chord, and, second, with chromatic tones mostly used for voice-leading and embellishment, i.e., as passing and neighboring tones or chromatic approaches. On the other hand, intrinsically non-diatonic chords (e.g., diminished chords) exhibit greater pitch flexibility, for instance by employing octatonic scales or interpreting them as chords with omitted tonic. Overall, chromaticism, voice-leading, figurations, and embellishments have been present since early jazz styles, thus generally blurring strict scale–chord associations. This suggests that a naïve prescriptive model of mapping chords to scales does not adequately describe actual practice. A more integrated view of vertical and horizontal pitch organization is necessary to fully understand (and master) jazz improvisation.

Folk Singing, Local Accent, and Generational Change in Jiuzhaigou

Xuhong Qiu (University of Cologne); Emilia Parada-Cabaleiro (Nuremberg University of Music)

In Chinese folk singing (民歌) and traditional opera (曲), melodic movement is often closely related to the tonal and prosodic features of language. Previous computational research on Beijing opera has shown that linguistic tone contours can shape the melodic characteristics of Chinese vocal genres, although the tone melody relationship may vary across repertories and dialectal contexts. While recent cross cultural studies have shown that songs often differ from speech through higher pitch, slower temporal rate, and greater pitch stability, this paper shifts the focus from universal speech song differences to a more local question: how does folk singing preserve, exaggerate, or reorganize prosodic features of local speech in a multilingual minority region increasingly shaped by Mandarin influence? This study investigates the relationship between local speech, Mandarin influence, and folk singing in the Jiuzhaigou region of Southwest China, with particular attention to age related variation. The project focuses on one selected local speech community in Jiuzhaigou and compares vocal materials from older tradition bearers and younger local speakers. Older participants will be asked to perform local folk songs, recite or speak song texts, and provide spoken narratives in the local speech variety and, where possible, in Mandarin. Younger participants, who may not actively know the songs, will provide spoken narratives and comparable speech materials in Mandarin and, where possible, in the local speech variety. This design treats younger speakers not as direct equivalents to older singers, but as a reference group for examining language shift, Mandarin influence, and changing local accent features. The analysis will use audio signal processing methods to examine pitch contour, temporal organization, syllable duration, speech rate, pitch stability, and selected prosodic features relevant to local accent. Particular attention will be paid to the relationship between lyric recitation and sung performance, including how linguistic prosody is aligned with or transformed by melodic structure. By combining acoustic and phonetic analysis with ethnomusicological attention to repertoire, transmission, and performance context, the study aims to show how Jiuzhaigou folk singing mediates between local speech, Mandarin influence, melodic form, and cultural memory.

Corpus Homogeneity and Musical Change Over a Century of Folk Song Collection in Ukrainian Folk Song

Anna Aljanaki (University of Music and Performing Arts Graz); Inna Lisniak (Estonian Literary Museum); Liudmyla Efremova (Institute of Art Studies, Folkloristics and Ethnology of the National Academy of Sciences)

A corpus of 1,805 symbolically encoded Ukrainian folk songs collected in the Podillia region spans a whole century of folk material collection between 1914 and 2018 (Lisniak, Aljanaki, and Efremova 2025) across three distinct historical periods: the early fieldwork stage (up to 1920), interwar (1920 - 1944), the stage of institutional archiving (1945–1990), and the contemporary stage (1991 onward). During these periods, Ukraine underwent profound political and social upheaval. In this research we examine the homogeneity of this corpus and melodic similarity across periods. First, we find that there is a profound shift in gender distribution (37.1% of male informants before the war to 3.8% after the war to 0% in contemporary period). Also, the thematic composition of collected songs changes substantially. Ballads shrink from making 54% of the early collection efforts to only 12%, most probably reflecting the Soviet suppression of ballad content, and a post-independence collecting agenda that prioritized ceremonial and ritual repertoire as most endangered. The opposite happens with wedding songs, which rise from 0% to 22%. We identify 57 songs whose catalogue number (encoding plot identity) appears in the corpus in three periods (before 1945, after 1945 and in contemporary period, obtaining 404 individual transcriptions. We extract the upper voice as the melody, and compute melodic similarity using weighted edit distance on interval sequences, where substitution cost is proportional to the size of the interval difference. We find that ritual songs (n=30) achieve a mean core-group WED similarity of 0.631, compared to 0.587 for non-ritual songs (n=25); Mann-Whitney U=483.0, p=0.035. This suggests that ritual songs preserve the overall melodic arc (the direction and approximate size of intervals) more consistently across periods than ballads. Outlier analysis reveals that 43 of 57 groups contain at least one melodic outlier. Comparing variants within the same period to those spanning different periods we find that for non-ritual songs, within-period similarity (mean WED 0.619) is significantly higher than cross-period similarity (0.532), Mann-Whitney $p < 0.01$. For ritual songs, within-period (0.573) and cross-period (0.575) similarity are virtually identical, $p = 0.5$. In other words, non-ritual/ballad melodies cluster within periods, ritual song melodies are preserved across periods. Ambitus steadily increases over the century of collection from a mean of 9.6 semitones in the early fieldwork period to 10.9 semitones in contemporary (Kruskal-Wallis H=36.22, $p < 0.0001$). Small ambitus is a widely recognized feature of archaic vocal traditions, where melodic range is often pentatonic or hexatonic, showing that more archaic performance tradition is slowly lost.

References

[1] Inna Lisniak, Anna Aljanaki, and Liudmyla Efremova. (2025) A corpus of Ukrainian folk songs from Podillia. The Proceedings of the 9th Digital Humanities in the Nordic and Baltic Countries Conference

Genre in a Phrase: Identifying the Formulaic Expressions That Distinguish German Popular Music Genres

Markus Radke (Technische Universität Berlin); Steffen Lepa (Technische Universität Berlin)

Lyrics are central to popular music, shaping emotional expression as well as conveying social meaning, and thus serve as important stylistic markers for popular music genre analysis. Formulaic expressions, including idioms, vocalizations, slang, and lexical single and multiword expressions, have proven valuable for identifying distinct lyrical genre styles in bilingual German popular music. Yet it remains unclear how many types of formulaic expressions are actually needed to differentiate genres and which are most important to characterize them. We address this through a secondary analysis of the bilingual POPTRAG corpus (N = 129,367 tracks), which comprises popular music frequently listened to in Germany, enriched with music genre labels from MusicBrainz condensed into 12 genre categories. We extract variable-length n-grams from English lyrics and translated German lyrics, then apply the "fightin' words" approach to identify a genre-discriminating subset. Structural topic modeling on this subset, with genre included as a prevalence covariate, is then used to uncover latent types of formulaic language; the optimal number of latent types is chosen using penalized held-out likelihood. Using type probabilities for each track, we then train machine-learning models for genre classification and assess performance and permutation-based feature importance on a held-out sample. Our results show that specific types of formulaic expressions function as distinctive genre markers. Although our analysis is limited to popular music in Germany, the approach provides a reproducible framework for comparative studies on other culturally-bounded popular music corpora in future research.

From Display to Instrument: An Experimental VR Reconstruction of the Golden Lyre of Ur

Yaozhong Zhang (University of Würzburg); Dahlia Shehata (University of Würzburg); Christof Weiß (University of Würzburg); Sebastian von Mammen (University of Würzburg)

The Golden Lyre of Ur (c. 2500 BCE) is one of the iconic musical instruments of ancient Mesopotamia. Its original strings and body had decayed long ago since its excavation, leaving only the metal cladding, so many open questions about the instrument—its string materials, tuning system, playing technique, and sound—remain disputed. Researchers have built several playable physical replicas from archaeological evidence, including plaster casts, photographs, and paintings. However, a physical reconstruction is inherently constrained. It is impractical for large-scale replication, sensitive to environmental conditions, and forces a single set of design choices that researchers cannot easily modify to explore other conjectures. We present an physics-based virtual reality (VR) system that reconstructs a playable lyre, allowing users to interact with it using natural hand gestures captured by optical hand tracking. In this virtual environment, we represent the instrument's geometric structure, string materials, and interaction mechanics as explicitly adjustable parameters. Guided by archaeomusicological hypotheses, we provisionally tuned the virtual lyre to a heptatonic scale and configured it with eleven gut strings. Users can individually retuned these strings in real time by rotating the pegs, allowing them to test and compare alternative tuning configurations during play. To compensate for the lack of haptic feedback in virtual interaction, we developed a visual feedback framework that conveys string contact and vibration through layered visual cues, enabling playing techniques such as plucking and muting. For music archaeologists, the system affords a new mode of engagement: an explainable, experimental reconstruction they can freely play and reconfigure between conjectures. Beyond research, it enables a broader public to play the instrument rather than observe it as a static object — an embodied encounter with a musical culture four thousand years old. We plan to present a poster with a live VR demo, inviting attendees to experience this firsthand.

Evaluating the Variety of Harmonic Qualia in Transformer-Based Artificially Generated Music

Gabriele Cecchetti (The MARCS Institute for Brain, Behaviour and Development); Roger Dean (The MARCS Institute for Brain, Behaviour and Development)

A core goal of artificial music-generation systems is to produce novel musical materials, often as a continuation of a user-defined seed. Artificial systems often exhibit suboptimal novelty-seeking behaviour compared to humans: for example, they may converge to uniform output (e.g., mode collapse), merely replicate the content of the seed, or diverge into unconstrained random variety (Dean & Forth, 2020). In this study, we investigate to what extent a transformer-based neural-network model replicates the novelty-seeking behaviour of human artists by comparing human and artificial continuations of identical seeds in terms of the variety they introduce. We hypothesise that, if neural-networks acquire novelty-seeking behaviour like that reflected in the training data, the artificial continuations might exhibit similar degrees of variety to the human continuations, distinct from randomly generated continuations obtained by sampling notes from the training data. We quantify variety with different metrics based on wavescapes (Viaccoz et al., 2022), which analyze a piece in terms of a hierarchical organization of the six harmonic qualia (e.g., diatonicity, triadicity, octatonicity) identified by decomposing pitch-class distributions through Discrete Fourier Transform: (1) the prevalence of a particular DFT component in the high hierarchical levels of the wavescape (global prototypicality after Laneve et al., 2023), (2) the diversity of qualia across the wavescape (fragmentation; Laneve et al., 2023), and (2) the size of the region of the Fourier Qualia Space (Pereira et al., 2024) traversed over the course of the piece. Non-parametric pairwise comparisons between human, artificial, and randomly generated continuations conducted over 5000 seeds drawn from the MAESTRO dataset (Hawthorne et al., 2018) consistently indicate that artificial continuations match the variety of real continuations more than randomly generated ones, across all metrics. These results suggest that neural-network-based generation captures to some extent the novelty-seeking behaviour reflected in human music in terms of how harmonic qualia are hierarchically organized.

References

[1] Dean, R. T., & Forth, J. (2020). Towards a Deep Improviser: A Prototype Deep Learning Post-Tonal Free Music Generator. Neural Computing and Applications, 32, 969–979.

[2] Hawthorne, C., Stasyuk, A., Roberts, A., Simon, I., Huang, C.-Z. A., Dieleman, S., Elsen, E., Engel, J., & Eck, D. (2018, October). Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset. ICLR.

[3] Laneve, S., Schaerf, L., Cecchetti, G., Hentschel, J., & Rohrmeier, M. A. (2023). The Diachronic Development of Debussy's Musical Style: A Corpus Study with Discrete Fourier Transform. Humanities and Social Sciences Communications, 10(1), 1–13. https://doi.org/10.1057/s41599-023-01796-7

[4] Pereira, S., Affatato, G., Bernardes, G., & Moss, F. C. (2024). Fourier Qualia Wavescapes: Hierarchical Analyses of Set Class Quality and Ambiguity. In T. Noll, M. Montiel, F. Gómez, O. C. Hamido, J. L. Besada, & J. O. Martins (Eds.), Mathematics and Computation in Music (Vol. 14639, pp. 317–329). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-60638-0_25

[5] Viaccoz, C., Harasim, D., Moss, F. C., & Rohrmeier, M. A. (2022). Wavescapes: A Visual Hierarchical Analysis of Tonality Using the Discrete Fourier Transform. Musicae

Scientiae, 10298649211034906. https://doi.org/10.1177/10298649211034906

Folk Music Lyrics Transcend Typological Boundaries Cross-Culturally (But It May Also Be the Typologies' Fault)

Anna Dvořáková (Charles University); Jan Hajič, jr. (Charles University)

Music is universally present in human societies [5]. Ethnomusicologists have long been documenting the diverse expressions of human musicality across social contexts, and comparative musicology has recently brought studies of folksong to a more global scale [8, 6]. However, such cross-cultural research has not yet been conducted on lyrics, even though for the construction of culture and shared identity through music, what people sing about in different social contexts may be as important as how they sing. We combined five folksong lyrics corpora with social context typologies (Dutch, Czech, Ukrainian, Estonian, Korean [7, 3, 4, 2]), machine-translated them to English (with evaluation by native speakers), and applied a pre-trained topic model [1] to obtain song representations in a shared space. Lyrical content is at best partially related to the social context of folksongs across all languages, although with considerable differences between individual contexts (scores ranging from 0.15 to 0.94) as well as between entire corpora (the worst-identifiable Korean labels and best-identifiable Czech labels are roughly equivalent, 50 to 60 %). Across combined wedding songs, separability by culture of origin ranges from 40 to 80 %. Czech and Ukrainian songs are, counter-intuitively, not predominantly confused for each other. Analysing where wedding songs fall in the topic space, we see a clearly separated Dutch cluster, a Korean control group (marriage laments) peppered with individual songs from other (non-Dutch) corpora, as well as topic-specific clusters with songs of several languages (e.g., Dutch, Korean, and Ukrainian for "fate"). Based on how the category of wedding songs is conceptualised in different corpora, and the high variance of per-corpus results, the next steps towards cross-cultural analysis of folksong lyrics should likely include re-examining the existing ethnomusicological typologies. These experiments, though early-stage, indicate that a previously unobserved web of cross-cultural relationships may be uncovered through the study of what people sing across the world's diverse folk musics.

References

[1] Maarten Grootendorst. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure". In: arXiv preprint arXiv:2203.05794 (2022). URL: https://doi.org/10.48550/arXiv.2203.05794.

[2] Danbinaerin Han, Rafael Caro Repetto, and Dasaem Jeong. Finding Tori: Self-supervised Learning for Analyzing Korean Folk Song. en. arXiv:2308.02249 [cs]. Aug. 2023. DOI: 10.48550/arXiv.2308.02249. URL: http://arxiv.org/abs/2308.02249 (visited on 10/05/2025).

[3] Matěj Kratochvíl, Petr Nuska, and Zdeněk Vejvoda. "Písňovna: Portál hudebního folkloru Čech, Moravy a Slezska a jeho využití v pedagogické praxi [Pisnovna: The Portal on the Folk Music of Bohemia, Moravia, and Silesia and Its Use in Educational Practice]." In: Digitální technologie v hudební pedagogice. Ed. by Milan Nedělka and Petra Bělohlávková. Praha : PedF UK, 2025, pp. 31–40. URL: https://publications.cuni.cz/bitstream/handle/20.500.14178/3269/Digitalnitechnologiedifinitivn%C3%AD.pdf.

[4] Inna Lisniak, Anna Aljanaki, and Liudmyla Efremova. "A Corpus of Ukrainian Folk Songs from Podillia". en. In: Digital Humanities in the Nordic and Baltic Countries Publications 7.2 (2025). ISSN: 2704-1441. DOI: 10 . 5617 / dhnbpub.12305. URL: https://journals.uio.no/dhnbpub/article/view/12305 (visited on 10/05/2025).

[5] Samuel A. Mehr et al. “Universality and diversity in human song”. In: Science 366.6468 (Nov. 2019). Publisher: American Association for the Advancement of Science, eaax0868. DOI: 10.1126/science.aax0868. URL: https://www.science.org/doi/10.1126/science.aax0868 (visited on 10/07/2025).
[6] Sam Passmore et al. “Global musical diversity is largely independent of linguistic and genetic histories”. In: Nature Communications 15.1 (May 2024). ISSN: 2041-1723. DOI: 10.1038/s41467-024-48113-7. URL: http://dx.doi.org/10.1038/s41467-024-48113-7.
[7] Peter Van Kranenburg, Martine De Bruin, and Anja Volk. “Documenting a song culture: the Dutch Song Database as a resource for musicological research”. In: International Journal on Digital Libraries 20.1 (2019), pp. 13–23.
[8] Anna L. C. Wood et al. “The Global Jukebox: A public database of performing arts and culture”. en. In: PLOS ONE 17.11 (Nov. 2022). Ed. by Steven R. Livingstone, e0275469. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0275469. URL: https://dx.plos.org/10.1371/journal.pone.0275469 (visited on 12/17/2025).

GRAP: A Greek Rap Dataset Towards Benchmarking Cross-Cultural Rap Analysis

Polyxeni Pouliou (Universitat Pompeu Fabra); Siddharth Saxena (Universitat Pompeu Fabra)

The development of varied datasets and automated tools remains a challenge in the MIR community. Depending on the target MIR task, researchers often rely on carefully curated datasets, the majority of which focus on Western music traditions, whether it is popular, or classical music. Drawn by our motivation to explore rap music, we noticed that, despite the growing interest in rap, the genre remains underrepresented in publicly available MIR resources. To the best of our knowledge, the only public rap-focused dataset is MCFlow (Condit-Schultz, 2016), which contains 124 English-language rap songs from the Billboard charts. To address this gap, we present our work GRAP, a Greek Rap Dataset, that will comprise 70 songs by influential Greek rap artists. Our dataset includes over 140 verses annotated with manually verified syllable-level timestamps, as well as segment-level boundaries. In addition to the annotations, we provide metadata files, YouTube source links, and alignment data for each track, following the reproducibility principles adopted in the Harmonix Set Dataset (Nieto et al., 2019). Our dataset is a work in progress and aims to initiate and support future research on rap music analysis, cross-cultural comparisons, and the evaluation of Automatic Speech Recognition (ASR) systems on rhythmically complex vocal performances.

Creation and Analysis of a Multimodal Film Music Dataset

Farshad Jafari (University of Cambridge); Claire Arthur (Georgia Institute of Technology)

Despite its cultural importance and more than a century of history, film music remains comparatively understudied. To help address this gap, we present a multimodal film music dataset sampled across 12 countries, 5 genres, and 3 decades, comprising 1,800 films and 633 hours of music aligned with English subtitles and visual features. Beyond the dataset itself, the project has two goals: first, to make the data accessible through a platform for exploration, visualization, and export of statistics; and second, to support reuse by releasing an open-source extraction framework that enables researchers to queue and monitor jobs, integrate extraction methods, and maintain versioning. The dataset includes features such as tempo, loudness variation, dialogue density, lexical density, shot boundaries, and dominant color, as well as embeddings and semantic annotations for music and image frames, including mood, genre, instrumentation, object type, size, and location. To assess annotation reliability, we conducted a listening study in which participants rated the presence of music, dialogue, and Foley, and evaluated system-suggested genre, instrument, and mood labels. Results indicate music segmentation and dialogue removal, moderate Foley removal, and alignment between human judgments and the model's top-ranked labels. At the corpus level, film music over the past three decades shows darker spectral profiles and reduced harmonic change density, suggesting a shift toward more sustained, texture-driven scoring. Cross-culturally, Indian film music stands out for high brightness and event density. Genre-specific patterns also emerge: horror cues favor electronic timbres, while action cues are associated with orchestral textures; action also exhibits the greatest loudness variation, whereas drama shows the lowest. Cross-modal analyses of music and narrative reveal tempo peaks near the 25%, 50%, and 75% points of a film, consistent with three-act screenplay structure. Together, these findings demonstrate the dataset's value for cross-cultural and cross-modal research on film music.

From Harmonic Grammar to Traceable Reharmonisation: A Rule-Based Framework for Jazz Harmony

George Konstandelakis (National and Kapodistrian University of Athens); Xristina Anagnostopoulou (National and Kapodistrian University of Athens)

Reharmonisation is a central practice in jazz performance, arranging, and composition, yet remains difficult to model computationally in ways that are both musically convincing and transparent to the user. The issue is visible in contemporary generative systems, where plausible harmonic outputs are often produced without transparent reasoning or meaningful user control. This paper makes a twofold contribution: it extends a jazz-harmonic grammar through a new set of reharmonisation rules and approaches, and implements this grammar in a rule-based framework for explainable, user-controlled reharmonisation. Reharmonisation is treated as a musical process involving functional reinterpretation, chromatic colour, stylistic choice, and voice-leading coherence. The user defines musically meaningful parameters, including a target level of harmonic complexity, locked chords, and other constraints, from which the system produces a stable harmonic description prior to generation. This remains fixed during evaluation and candidate selection, making each stage of the pipeline traceable. The system builds on a harmonic grammar of 11 rules (21 with subrules) originally developed for jazz piano improvisation analysis, extending it into a knowledge-representation hierarchy of rules, approaches, harmonic families, and stylistic classes. These formalise jazz-harmonic procedures across functional, chromatic, modal, bass-driven, and symmetric territories, organised into seven complexity-graded families. A user-defined Complexity Target controls family activation, weighting, exploratory behaviour, and harmonic-regime selection. Candidate reharmonisations are generated according to musical context, stylistic constraints, and voice-leading requirements, rendered in staff notation, and evaluated through a Target Vector model with three interpretable dimensions: surface chromatic colour, functional change, and structural intervention. Each session presents multiple candidates with derivation traces, rule paths, active family weights, and explanatory annotations. The work contributes to music knowledge representation, knowledge-based music generation, computational creativity, and explainable models of harmonic practice.

References

[1] M. A. Boden, The Creative Mind: Myths and Mechanisms, 2nd ed. London: Routledge, 2004.

[2] G. A. Wiggins, "A Preliminary Framework for Description, Analysis and Comparison of Creative Systems," Knowledge-Based Systems, vol. 19, no. 7, pp. 449–458, 2006.

[3] D. Herremans, C.-H. Chuan, and E. Chew, "A Functional Taxonomy of Music Generation Systems," ACM Computing Surveys, vol. 50, no. 5, Article 69, pp. 1–30, 2017.

[4] D. Cope, Experiments in Musical Intelligence. Madison, WI: A-R Editions, 1996.

[5] M. J. Steedman, "A Generative Grammar for Jazz Chord Sequences," Music Perception, vol. 2, no. 1, pp. 52–77, 1984.

[6] M. Rohrmeier, "Towards a Generative Syntax of Tonal Harmony," Journal of Mathematics and Music, vol. 5, no. 1, pp. 35–53, 2011.

[7] M. Rohrmeier, "The Syntax of Jazz Harmony: Diatonic Tonality, Phrase Structure, and Form," Music Theory and Analysis, vol. 7, no. 1, pp. 1–63, 2020.

[8] D. Harasim, M. Rohrmeier, and T. J. O'Donnell, "A Generalized Parsing Framework for Generative Models of Harmonic Syntax," in Proc. ISMIR, 2018.

[9] D. Harasim, T. J. O'Donnell, and M. Rohrmeier, "Harmonic Syntax in Time: Rhythm Improves Grammatical Models of Harmony," in Proc. ISMIR, 2019.
[10] D. Harasim, C. Finkensiep, P. Ericson, T. J. O'Donnell, and M. Rohrmeier, "The Jazz Harmony Treebank," in Proc. ISMIR, 2020.
[11] F. Lerdahl and R. Jackendoff, A Generative Theory of Tonal Music. Cambridge, MA: MIT Press, 1983.
[12] P. Berliner, Thinking in Jazz: The Infinite Art of Improvisation. University of Chicago Press, 1994.
[13] H. Crook, How to Improvise: An Approach to Practicing Improvisation. Advance Music, 1991.
[14] J. Gillick, K. Tang, and M. R. Keller, "Learning Jazz Grammars," in Proc. Sound and Music Computing (SMC), 2009.
[15] M. Kaliakatsos-Papakostas, K. Velenis, L. Pasias, C. Alexandraki, and E. Cambouropoulos, "An HMM-Based Approach for Cross-Harmonization of Jazz Standards," Applied Sciences, vol. 13, no. 3, 1338, 2023.
[16] R. Rawlins and N. E. Bahha, Jazzology: The Encyclopedia of Jazz Theory for All Musicians. Hal Leonard, 2005.
[17] D. Berkman, The Jazz Harmony Book. Sher Music Co., 2013.
[18] R. Felts, Reharmonization Techniques. Berklee Press / Hal Leonard, 2002.
[19] B. Ligon, Jazz Theory Resources, Vol. 2. Hal Leonard, 2001.
[20] D. Liebman, A Chromatic Approach to Jazz Harmony and Melody. Advance Music, 2015.
[21] T. Pease and K. Pullig, Modern Jazz Voicings. Berklee Press / Hal Leonard, 2001.
[22] B. Boyd, Jazz Chord Progressions. Hal Leonard, 1997.
[23] M. Levine, The Jazz Theory Book. Sher Music, 1995.
[24] J. Mulholland and T. Hojnacki, The Berklee Book of Jazz Harmony. Berklee Press, 2013.

Bridging Harmonic Representations Through a Multi-Modal Encoding

Patricia Garcia-Iasci (Universidad de Alicante. Universidad de Salamanca); Johannes Hentschel (Anton Bruckner University); Fabian C. Moss (University of Würzburg); David Rizo (University of Alicante, Instituto Superior de Enseñanzas Artísticas de la Comunidad Valenciana)

Advancements in Computational Musicology are contingent upon the size and quality of the available research data. A large number of formal harmony encoding schemes exist, yet reconciliation of harmony labels adhering to disparate standards remains a difficult challenge. Although proposals of unified models exist (e.g., Hentschel et al., 2022), none have been implemented, compelling researchers and projects to define custom mappings and ad-hoc workarounds. One of the main challenges lies in the diversity of projects using different harmonic representations for their own purposes (Moss et al., 2019; Tymoczko et al, 2019; Harasim et al., 2020; López et al., 2020; Ju, 2021; Gotham et al., 2023; Hentschel et al., 2025). Sometimes, even with a bespoke protocol that combines several of them to suit all analysis needs (Rizo et al., 2015). While this is an effective solution in the short term, it could entail information loss and may require allocating considerable time and resources to clean and adapt those data structures and their harmonic meanings for specific tasks (Hentschel et al., 2026; Martínez-Sevilla et al., 2025). In addition to the complexity involved, the resulting pragmatic data models are typically project-specific, which undermines their interoperability and limits the scalability and transferability of the computational approaches they would otherwise enable. Here, we present HAMON, a grammar-based approach for standardising the encoding of harmony annotations as a universal exchange layer for musical notation files. It currently aligns 10 different formats (MEI, MusicXML, Humdrum, LilyPond, ABC, MuseScore, DCML, Harte, iReal Pro, music21) into a typed Abstract Syntax Tree (AST) according to a formal grammar, with reference implementations in TypeScript and Python available. This proposal is designed for multimodal use cases and allows the encoding of different sources to facilitate musical research across time, periods, and genres.

References

[1] Gotham, M., Micchi, G., López, N. N., y Sailor, M. (2023). When in Rome: A Meta-corpus of Functional Harmony. Transactions of the International Society for Music Information Retrieval, 6(1), 150–166. https://doi.org/10.5334/tismir.165

[2] Harasim, D., Finkensiep, C., Ericson, P., O'Donnell, T. J., & Rohrmeier, M. (2020). The jazz harmony treebank. In 21st International Society for Music Information Retrieval Conference (ISMIR, 2020) (pp. 207-215).

[3] Hentschel, J., Moss, F. C., McLeod, A., Neuwirth, M., & Rohrmeier, M. (2022). Towards a Unified Model of Chords in Western Harmony. In S. Münnich & D. Rizo (Eds), Music Encoding Conference Proceedings 2021 (pp. 143–149). https://doi.org/10.17613/5rxtc-wcj65

[4] Hentschel, J., Rammos, Y., Neuwirth, M., & Rohrmeier, M. (2025). A corpus and a modular infrastructure for the empirical study of (an)notated music. Scientific Data, 12(1), 685. https://doi.org/10.1038/s41597-025-04976-z

[5] Hentschel, J., Karystinaios, E., Widmer, G., & Neuwirth, M. (2026, May 28). Dilemmadata: On the interoperability of heterogeneous Roman numeral datasets [Music Encoding Conference]. MEC2026, Tokyo, Japan.

[6] López, N. N., Feisthauer, L., Levé, F., & Fujinaga, I. (2020). On local keys, modulations,

and tonicizations. In Digital Libraries for Musicology (DLfM 2020).
[7] Martinez-Sevilla, J. C., Foscarin, F., Garcia-Iasci, P., Rizo, D., Calvo-Zaragoza, J., & Widmer, G. (2025). Optical Music Recognition of Jazz Lead Sheets. In 26st International Society for Music Information Retrieval Conference (ISMIR, 2025).
[8] Moss, F., Neuwirth, M., Harasim, D., Rohrmeier, M., (2019) Statistical characteristics of tonal harmony: A corpus study of Beethoven's string quartets. PLOS ONE 14(6): e0217242. https://doi.org/10.1371/journal.pone.0217242
[9] Rizo, D., Illescas, P. R., & Inesta, J. M. (2015). Interactive melodic analysis. In Computational Music Analysis (pp. 191-219). Cham: Springer International Publishing.
[10] Tymoczko, D., Gotham, M., Cuthbert, M. S., & Ariza, C. (2019). The RomanText Format: A Flexible and Standard Method for Representing Roman Numeral Analyses. Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR, 123–129.

Poster Session C

Style Analysis of Western Classical Music Recordings with Robust Pitch-Class Features

Benjamin Henzel (University of Würzburg); Christof Weiß (University of Würzburg)

The formation of chords and chord progressions using pitches (or pitch classes) is a fundamental concept that has been present throughout a large part of the history of Western classical music. Their concrete realization is a determining characteristic of compositional style and has evolved over centuries. To examine such stylistic evolution based on audio recordings, it is necessary to extract features that are style-relevant and interpretable, but also robust towards different types of instrumentation. In this study, we employ an improved deep-learning-based approach that has proven to be particularly robust towards these conditions. Using this method, we compute pitch-class representations for the Cross-Era dataset, which consists of 1,600 recordings of Western instrumental music from four style eras. To perform style classification, we derive three types of tonal features from the computed pitch-class representations: tonal complexity measures, interval and chord type saliences, and chord transition probabilities. In our experiments, we find that style classification benefits from deep pitch-class representations as compared to traditional signal-processing features. In a second case study, we combine the Cross-Era dataset with the Carus Audio Corpus, resulting in a large corpus comprising 7.729 tracks of both instrumental and vocal music from approximately 450 years of Western classical music. Based on the deep pitch-class representations, we compute tonal features and investigate the stylistic evolution over time. The results confirm that robust tonal features enable meaningful analyses across different instrumentations and musical settings. Overall, our findings highlight the value of robust pitch-class representations for style analysis, leveraging the potential of deep learning while preserving interpretability of the tonal features.

Reconstructing 13th-Century Motets: A Comparative Study of Rule-Based and Deep Learning Models

Xuhong Qiu (University of Cologne); Emilia Parada-Cabaleiro (Nuremberg University of Music); Frank Hentschel (University of Cologne)

The 13th-century motet is one of the central genres of early polyphony, yet its internal compositional logic remains only partially understood. Although previous scholarship has examined its origins, notation, transmission, and individual structural features, there is still no generally accepted model explaining whether its contrapuntal organisation, vertical sonorities, and rhythmic structures follow recurrent principles across the repertory. This project addresses this problem by combining historical musicology with computational modelling and generative reconstruction. Building on the harMo13 project at the University of Cologne, the study uses a corpus of 73 three voice and four voice motets from major 13th-century sources (Montpellier Codex, the Bamberg Codex, La Clayette, and the Turin manuscript). The corpus will be standardised for symbolic analysis, including rhythmic normalisation and formats suitable for computational processing. Analytical work will focus on melodic profiles, intervallic combinations, sonority types, cadence patterns, rhythmic designs, voice activity, and larger structural relations. These features will be extracted and examined through symbolic analysis tools such as music21, Humdrum, and jSymbolic, as well as unsupervised machine learning methods such as clustering and dimensionality reduction. The central methodological aim is to compare two paradigms of musical generation: rule based modelling and data driven deep learning. Rule-based systems such as OpenMusic allow explicitly formulated hypotheses about motet structure to be tested, revised, or rejected. Deep learning models, including transformer based approaches and adapted symbolic music models, may reveal stylistic regularities that are difficult to formalise in advance. The project therefore treats generation not as autonomous composition, but as a heuristic method for historical inquiry. By comparing the successes and failures of these modelling approaches, the dissertation seeks to clarify which aspects of 13th-century motet organisation can be reconstructed explicitly, which emerge through corpus based learning, and which remain resistant to computational modelling. This contributes to medieval musicology and computational approaches to early polyphony.

Mixing Real and Synthetic Data for Handwritten Optical Music Recognition: Domain Adaptation on Debussy Manuscripts

Hugo Scheithauer (Inria); Thibault Clérice (Inria); Gonzalo Romero-García (EPITA)

Optical Music Recognition (OMR) converts score images into machine-readable symbolic formats [2]. Recent end-to-end systems report strong performance on benchmark datasets built from synthetic Common Western Music Notation (CWMN) material [4–6,9]. However, these benchmarks struggle to fully represent archival conditions: they do not capture the visual variability, degradation, and notation-specific difficulties found in handwritten historical manuscripts preserved in library collections [10]. We study data-efficient adaptation to a corpus of Claude Debussy score manuscripts held at the French National Library, using the Sheet Music Transformer (SMT) [5, 7] pretrained on the GrandStaff dataset [1, 8]. Our objective is to quantify the benefit of mixing synthetic and real documents for handwritten OMR within a single-composer, single-handwriting corpus. Our contribution is a two-step analysis of a mixed training setup. First, we fine-tune the SMT with real Debussy data and Smashcima (a handwritten music synthesis framework that renders MusicXML files into realistic manuscript-like images. [3]) synthetic pages generated from Debussy MusicXML transcriptions under diverse handwriting styles. Second, we evaluate style-matched synthesis by combining real Debussy samples with Smashcima pages rendered in Debussy-like handwriting. We compare against zero-shot and real-only baselines to isolate gains from synthetic augmentation and style alignment. To prevent data leakage, synthetic generation is restricted to training folds, and all conditions share identical splits, model backbone, and evaluation metrics (symbol, character, and line error rates). This design isolates three effects: the manuscript domain gap, the impact of synthetic augmentation beyond real-only fine-tuning, and the benefit of handwriting-style matching over generic synthetic variation. More broadly, we provide an empirical basis for combining real and synthetic supervision when adapting pretrained OMR models to historical collections.

References

[1] Jorge Calvo-Zaragoza and David Rizo. Camera-PrIMuS: Neural End-to-End Optical Music Recognition on Realistic Monophonic Scores. In Emilia Gómez, Xiao Hu, Eric Humphrey, and Emmanouil Benetos, editors, Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018, Paris, France, September 23-27, 2018, pages 248–255, 2018.

[2] Francisco J. Castellanos, Antonio Javier Gallego, and Ichiro Fujinaga. Deep Learning for Optical Music Recognition: A Review. Authorea Preprints, February 2025.

[3] Jiří Mayer, Pavel Pecina, and Jan Hajič jr. Smashcima: Full-Page Handwritten Music Document Synthesizer. In Proceedings of the 12th International Conference on Digital Libraries for Musicology, DLfM '25, pages 119–123, New York, NY, USA, September 2025. Association for Computing Machinery

[4] Jiří Mayer, Milan Straka, Jan Hajič jr., and Pavel Pecina. Practical End-to-End Optical Music Recognition for Pianoform Music. volume 14809, pages 55–73. 2024. arXiv:2403.13763 [cs].

[5] Antonio Ríos-Vila, Jorge Calvo-Zaragoza, and Thierry Paquet. Sheet Music Transformer: End-To-End Optical Music Recognition Beyond Monophonic Transcription, April 2024. arXiv:2402.07596 [cs].

[6] Antonio Ríos-Vila, Jorge Calvo-Zaragoza, David Rizo, and Thierry Paquet. End-toEnd

Full-Page Optical Music Recognition for Pianoform Sheet Music, September 2024. arXiv:2405.12105 [cs].
[7] Antonio Ríos-Vila, Jorge Calvo-Zaragoza, David Rizo, and Thierry Paquet. Sheet Music Transformer ++: End-to-End Full-Page Optical Music Recognition for Pianoform Sheet Music, May 2024. arXiv:2405.12105 [cs] version: 2.
[8] Antonio Ríos-Vila, Jose M Íñesta, and Jorge Calvo-Zaragoza. End-to-End Full-Page Optical Music Recognition for Mensural Notation. In Preeti Rao, Hema Murthy, Ajay Srinivasamurthy, Rachel Bittner, Rafael Caro Repetto, Masataka Goto, Xavier Serra, and Marius Miron, editors, Proceedings of the 23rd International Society for Music Information Retrieval Conference, ISMIR 2022, Bengaluru, India, December 4-8, 2022, pages 226–232, 2022.
[9] Antonio Ríos-Vila, David Rizo, Jose M Íñesta, and Jorge Calvo-Zaragoza. End-to-end optical music recognition for pianoform sheet music. International Journal on Document Analysis and Recognition (IJDAR), 26(3):347–362, September 2023.
[10] Lukas Tuggener, Raphael Emberger, Adhiraj Ghosh, Pascal Sager, Yvan Putra Satyawan, Javier Montoya, Simon Goldschagg, Florian Seibold, Urs Gut, Philipp Ackermann, Jurgen Schmidhuber, and Thilo Stadelmann. Real World Music Object Recognition. Transactions of the International Society for Music Information Retrieval, 7(1), January 2024.

Mining Contrast Patterns in the Jazz Guitar Improvisations: Comparing Pat Metheny with Five Post-bop Guitarists

Arne Eigenfeldt (Simon Fraser University); Darrell Conklin (University of the Basque Country; UPV/EHU)

Contrast pattern mining is the process of finding patterns whose frequency differs significantly from one group of data to another (Conklin, 2024). This paper uses contrast pattern mining to investigate how the improvised melodic interval patterns of progressive jazz guitarist Pat Metheny (b. 1954) differ from those of guitarists from the post-bop period, from the mid-1950s to the mid-1970s. For this, we used transcriptions from 29 of Metheny's solos; these transcriptions span five decades – from the 1976 album Bright Size Life through his Unity Band's 2012 album. These solos were contrasted with five jazz guitarists from the post-bop period – Wes Montgomery, Jim Hall, Grant Green, Kenny Burrell, and Barney Kessel; these solos were recorded between 1954 and 1977. When comparing Metheny's patterns to those found in non-guitarists in jazz from previous research (Conklin and Eigenfeldt, 2026), we found that there were Metheny-specific patterns that appeared also in other instruments. These include alternating major seconds or diatonic scalar runs, both of which fall easily under the performer's hand on the instrument. However, we found unique patterns in Metheny's improvisations, stemming from how Metheny exploits the guitar's ability to play a pattern and easily move the pattern up or down the fingerboard, thereby producing parallel intervals and patterns at non-diatonic (i.e., chromatic) levels. Despite Metheny's music remaining fundamentally diatonic and tonal, this chromatic shifting of patterns results in a greater degree of chromaticism in his over-represented patterns when compared to the post-bop players. Noticeably, some diatonic patterns appear only in Metheny's improvisations: for example, a six-note ascending minor scale pattern occurs only in Metheny's solos and not in the post-bop players, as does the ascending major triad in second inversion. Throughout this research, we make a distinction between statistical significance versus deep musical significance; the above two examples of Metheny's diatonic patterns are statistically significant yet difficult to attribute any musical significance to. In every case, we look at the context of the patterns within the improvisations, particularly noting how both groups juxtapose and possibly interlock patterns. This presentation will describe the corpus of post-bop and Metheny solos, outline the methods used for pattern discovery, and highlight some significant patterns, including some that we find musically interesting.

References

[1] Conklin, D. (2024) Mining Significant Sequential Contrast Patterns. In: Noll, T., Montiel, M., Gómez, F., Hamido, O.C., Besada, J.L., Martins, J.O. (eds.) Mathematics and Computation in Music. Pp. 387–392. Springer Nature Switzerland, Cham (2024)

[2] Conklin, D. and Eigenfeldt, A. (2026). Mining contrast patterns in jazz solos. In Machine Learning and Principles and Practice of Knowledge Discovery in Databases. ECML PKDD 2024. Communications in Computer and Information Science, vol. 2559, pages 402-408.

Musical Enculturation and Emotional Perception: Cross-Cultural Responses to Heptatonic and Pentatonic Melodies

Ryan Lu (Trinity School)

Music cognition research has demonstrated that cultural exposure shapes the perception and interpretation of musical structures. However, relatively few studies have directly examined emotional responses to pentatonic and heptatonic melodies across culturally diverse listeners. This study investigates how musical enculturation shapes emotional responses to pentatonic and heptatonic melodies. A total of 160 participants from 35 countries and regions were recruited through Prolific, an online platform to recruit research participants. Participants were assigned to two groups (Western familiar group and Eastern familiar group) based on self-reported familiarity with Western and Eastern cultures, enabling comparisons between culturally familiar and unfamiliar listeners. Participants listened to five short melodic stimuli, including two pentatonic, two heptatonic, and one mixed-scale excerpt, and evaluated each stimulus on six perceived emotional dimensions, including pleasure, calmness, sadness, nostalgia, tension, and excitement on 10-point scales ranging from 1 to 10. Analyses of variance (ANOVA) were conducted to statistically examine differences in emotional ratings across cultural familiarity groups and musical scale types. Preliminary analyses showed significant cross-cultural differences in emotional responses to both heptatonic and pentatonic melodies. For the heptatonic stimuli, significant group differences were observed primarily in excitement ratings ($p < .05$). For the pentatonic stimuli, significant differences emerged in calmness, pleasure, and nostalgia ratings, with one pentatonic melody showing particularly strong effects in nostalgia ($p < .001$). In contrast, mixed-scale stimuli showed no significant differences between groups across any emotional dimension. These findings suggest that the listeners' cultural familiarity influences the emotional interpretation of culture-specific tonal systems. The study contributes to a broader understanding of how culture shapes music cognition and informs the design of culturally responsive applications in music therapy, education, and recommendation systems.

Why Random Music Works: A Pitch-Class Space Account of Coherence in Mozart's Musikalisches Würfelspiel

Benjamin Oh (Lynbrook High School); Hee Yun Kim (Threaded Resonances)

In the era of AI-generated music, Mozart's Musikalisches Würfelspiel (K. 516f) raises a fundamental question: what makes algorithmic compositions sound coherent? How can random combinations of musical fragments remain stylistically convincing? We propose a two-factor explanation grounded in pitch-class space, offering an alternative to prevalent accounts emphasizing smooth voice leading as a primary source of coherence. We compare Mozart's 176 Menuet measures with three genre-matched corpora — Haydn, (3,119 measures), Beethoven (968 measures), and Schubert (1,547 measures) — spanning 3/4-meter writing, each represented as a normalised 12-dimensional pitch-class histogram. First, 88.6% of Mozart's Menuet measures cluster within a narrow D major tonal region (σ = 0.822), compared to flat distributions in all three corpora (σ = 1.096–1.507), all $p < 0.001$. This tonal homogeneity functions as a floor: any combination remains within a single sound world. Second, Mozart's measures exhibit the widest registral range among the four corpora (raw voice-leading distance 2.143 vs. 0.829–1.632), ensuring that each combination sounds distinct. These two properties, tonal constraint and registral freedom, create deceptive variety between the fragments. Similarly, 92.7% of Trio measures cluster in G major — the subdominant of D major — with structural uniformity virtually identical to the Menuet. No comparison corpus maintains tonal homogeneity from Menuet to Trio. A within-composer analysis of 16 Mozart string quartet Menuettos yields σ = 1.161–1.177 — significantly higher than the Dice Game and closer to that of Haydn, confirming that tonal homogeneity is specific to K. 516f. These findings offer a new account of musical coherence, suggesting it arises not from local voice-leading continuity but from global constraints in pitch-class space, and point to an alternative to probabilistic generative models: rather than assembling sequences from likely events, define a constrained space within which variation can freely occur.

Modelling Perceived Tension Using Hierarchical Harmonic Structure

Edward Hall (EPFL); Martin Rohrmeier (EPFL)

Harmonic structure plays an important role in the perception of musical tension. We propose a computational cognitive model of harmonic tension, based on the hypothesis that tension is created by the non-adjacent relationships between harmonies in hierarchical harmonic trees. Following Rohrmeier (2011), we use two classes of relationship: (1) prolongations, where the same harmonic region is maintained; and (2) preparations, where one chord creates an anticipation of another. We suggest that both can contribute to perceived tension—where tension gradually increases as a harmony is prolonged, and a preparing chord creates tension that is released at its resolution. We report a behavioural experiment that aimed to test perception of tension and can be used to fit and evaluate the proposed model. Stimuli consisted of keyboard works by Bach, Mozart, Schubert, Schumann, Mendelssohn, Chopin, and Liszt (duration M = 162 s). Using a slider, 29 participants provided continuous ratings of tension for each work. Works were annotated to identify chord labels and timings. Hierarchical tree structures were analysed following Rohrmeier (2011). A computational model was implemented that combined tension contributions for each span in a given tree—i.e., either the time between a preparing chord and its resolution, or between the start and end of a prolongation. For both categories of span, a tension profile was defined. Profile contributions were summed across the piece and profile parameters were fitted so as to minimise the sum-of-squares error between simulated tension and experiment data. Simulated tension predicted observed tension ratings for stimuli. Both prolongation and preparation spans contributed to perceived tension, but with different characteristics—as shown by their respective optimised parameters; tension from preparation decayed faster. Spans originating further from the surface had a greater effect on tension. The performance of the presented model is compared to existing models of tension.

Musical Form as a Projection of Drama: Parameter Progressions in the Wanderer–Mime Scene from Wagner's Ring

Pascal Schmolenzky (International Audio Laboratories Erlangen); Rainer Kleinertz (HfM Saar); Stephanie Klauk (Saarland University); Meinard Müller (International Audio Laboratories Erlangen); Christof Weiß (University of Würzburg)

Richard Wagner's music dramas represent a significant challenge for musicological analysis since they lack typical form-articulating features, such as clear caesuras, closing formulas, or architectonic block structures. Despite many promising approaches to the analysis of form and structure, systematic investigations of entire scenes, acts, or works remain comparatively rare. One such approach is Anthony Newcomb's analysis of Siegfried, Act III, Scene 1, which assumes a close correlation between dramatic and musical form observable through trajectories of time-dependent parameters such as tonality, tempo, and dramatic interaction (Newcomb, 1981). Building on this perspective, we use automated methods for the visualization of harmonic trajectories from audio recordings (Weiß, Kleinertz, Klauk, & Müller, 2026) and annotations from the Wagner Ring Dataset (Weiß et al., 2023) to derive a reduced representation that remains musically meaningful in this sense. As already shown by Schmolenzky, Klauk, Weiß, Kleinertz, and Müller (2025), Wagner appears to apply coherent strategies for parameter progressions analogous to dramatic events. Using the Wanderer–Mime scene, the second scene from Siegfried Act I, as a case study, we combine time-varying diatonic scale distributions, median tempo trajectories across 16 recordings, and annotations of vocal interaction on a shared musical time axis. Although the scene belongs to the most frequently discussed passages of the Ring, the resulting multi-layer visualization, by providing a synoptic overview of large spans of music, facilitates a deeper understanding of the scene's formal organization and opens new perspectives on its relation to large-scale formal processes.

References

[1] Newcomb, A. (1981). The birth of music out of the spirit of drama: An essay in Wagnerian formal analysis. 19th-Century Music, 5 (1), 38–66.

[2] Schmolenzky, P., Klauk, S., Weiß, C., Kleinertz, R., & Müller, M. (2025). A multidimensional approach to opera analysis: Harmony, tempo and dramatic interaction in Wagner's 'Siegfried'. In Proceedings of the international society for music information retrieval conference (ISMIR). Daejeon, Korea.

[3] Weiß, C., Arifi-Müller, V., Krause, M., Zalkow, F., Klauk, S., Kleinertz, R., & Müller, M. (2023). Wagner Ring Dataset: A complex opera scenario for music processing and computational musicology. Transaction of the International Society for Music Information Retrieval (TISMIR), 6 (1), 135–149.

[4] Weiß, C., Kleinertz, R., Klauk, S., & Müller, M. (2026). Computational strategies for measuring and visualizing tonal structures: Towards a large-scale analysis of Wagner's 'Ring' cycle. ACM Journal on Computing and Cultural Heritage (JOCCH).

Music Score Comparison Based on a Tree-Structured Intermediate Representation

Florent Jacquemard (Inria); Xiaobo Wang (INALCO)

We propose a new algorithm for comparing digital music scores in various encodings (e.g. MusicXML or MEI). It is based on an intermediate abstract representation structured as labeled directed acyclic graphs (DAGs). Inspired by rhythm trees [4], these DAGs encode the metric and rhythmic structure of a score. Given two scores, after conversion into DAGs, our procedure calculates jointly: - a list of differences — localised by time positions in the input scores; - a time distance — cumulated duration of time intervals where scores differ; - an edit distance based on attributes of notational elements not related to time, called surface distance. The two later measures are normalised and combined into a similarity distance between scores. The DAG's nodes are labelled with basic events (notes, rests, ...) for leaves, and, for inner nodes, by operators acting on time intervals, like splitting into sub-intervals of equal duration (in a tuplet), or one bar and the remaining, or concurrent voice creation. Node sharing represent composed events such as dotted notes. This way, every node corresponds to a time interval. The comparison proceeds first bar-wise, aligning equal bars, then by a recursive parallel descent into sub-DAGs representing differing bars, stopping as soon as labels diverge. This yields linear time complexity and provides more explicit feedback than the ins/del/replace operations of classical edit distances, with ad-hoc difference codes associated to each pair of node labels. The exact list of codes can be adapted to use-cases. This approach can be applied e.g. to comparative corpus studies in musicology [5], or versioning in the context of score edition or preparation of corpora. As a proof of concept, we propose [1] two demos based on a corpus of piano sonata [6], and transcriptions of monophonic soli on the same jazz standard by different performers [7].

References

[1] `https://tsm-diff-00a041.gitlab.io`

[2] Foscarin, F., Jacquemard, F., & Fournier-S'niehotta, R. (2019). A diff procedure for music score files. Proceedings of the 6th International Conference on Digital Libraries for Musicology.

[3] Martinez-Sevilla, J. C., Cerveto-Serrano, J., Luna-Barahona, N., Chapman, G., Sapp, C., Rizo, D., & Calvo-Zaragoza, J. (2025). Sheet Music Benchmark: Standardized Optical Music Recognition Evaluation. Proceedings of the 26th International Society for Music Information Retrieval Conference.

[4] Agon, C., Haddad, K., & Assayag, G. (2002). Representation and rendering of rhythm structures. Proceedings of the Second International Conference on Web Delivering of Music. IEEE.

[5] Nachtwey, A., Moss, F. C., & Plaksin, A. V. K. (2025). Beyond Bars: Distribution of Edit Operations in Historical Prints. Music Encoding Conference (poster).

[6] Hentschel, J., Neuwirth, M., & Rohrmeier, M. (2021). The Annotated Mozart Sonatas: Score, harmony, and cadence. Transactions of the International Society for Music Information Retrieval, 4(1).

[7] Foster, D., & Dixon, S. (2021). Filosax: A dataset of annotated jazz saxophone recordings. Proceedings of the 22nd International Society for Music Information Retrieval Conference.

Just Intonation in Tonal Thought: Towards an Algorithmic Corpus Study

Jan-Martin Gebert (Columbia University)

It is widely acknowledged that the realization of Just Intonation (JI) in common-practice music is generally not possible. As a purely theoretical framework, however, JI may nevertheless help to illuminate aspects of tonal thought. The central issue in reconciling JI with tonal music is the comma problem. This arises when the same notated pitch is derived along different harmonic paths in JI, leading to conflicting simultaneous pitch identities (see Dahlhaus 1968). Rather than treating such pitch discrepancies, or commas, as negligible acoustic artifacts, I focus on situations in which they appear to be coherently introduced and subsequently resolved. The theorists Karg-Elert (1930; see Byrne 2018) and Hohlfeld (1989, 1994, 2003, n.d.) describe such instances as "comma-free modulations": progressions in which a comma introduced along one harmonic path is counterbalanced by a reciprocal path, leaving no residual comma. Roughly speaking, a comma-free modulation describes the rare case in which the comma problem is avoided for a relatively long span, making the passage JI-compatible. Two case studies, Beethoven's Sonata Op. 49 No. 2 and Coltrane's "Giant Steps," illustrate such comma-free modulations. A "miniature" corpus study of Mozart's piano sonatas suggests that they may occur with some regularity. The examples are presented through Tonnetz animations alongside audio tracks rendered in pure JI. Building on these base cases, I propose an algorithm for detecting JI-compatible passages in MEI/XML-encoded music, including microtonal MIDI export (see Ratib 2026). The algorithm is loosely based on Euler's consonance formulas, which were already used in attempts, initiated by JI apologist Martin Vogel, to create computerized JI instruments (Euler 1739; Vogel 1984; see also the project "Mutabor" in Schlemmer 2018). By examining possible associations between comma-free passages and compositional structure, the project ultimately aims to explore the relationship between JI and compositional decision-making in tonal music.

R3MO: A Multi-modal Dataset for Emotion Analysis in Orchestral Music

Joann Ching (Johannes Kepler University Linz); Carlos Eduardo Cancino-Chacón (Johannes Kepler University Linz); Gerhard Widmer (Johannes Kepler University Linz)

The R3MO Dataset is a multi-modal dataset with emotion annotations of Rachmaninoff's Piano Concerto No. 3 in D minor, Op. 30. It comprises 11 recordings of the concerto by different pianists and orchestras, a two-piano reduction of the orchestral score, and non-overlapping segment-wise emotion annotations. The segments are created according to the musical structure. Each segment is annotated by, on average, 8.9 students in a CS/AI course, with both dimensional (valence-arousal) and categorical descriptors (GEMS and GEMMES) on a 5-point Likert scale. While Russell's Circumplex [1] provides a general psychological model of affect in a two-dimensional space, the Geneva Emotional Music Scale (GEMS-9) [2] is a music-specific taxonomy designed to capture more nuanced, music-evoked emotions. Furthermore, the Geneva Musical Metaphors Scale (GEMMES) [3] extends this approach through metaphor-based descriptors, allowing intuitive labeling by annotators regardless of musical expertise. A set of low- and mid-level acoustic features extracted from the audio recordings is also provided to complement the annotations. To evaluate annotation reliability, we compute Krippendorff's alpha across emotional measures and quantify inter-annotator variability with the coefficient of variation. We further analyze relationships between categorical and dimensional annotations by correlating the categorical descriptors with valence–arousal ratings, mapping each descriptor onto the circumplex. Despite expected differences in inter-annotator agreement across these emotional measures, the resulting mapping shows consistency with prior neurological findings [4], which classify GEMS descriptors into high- and low- valence–arousal quadrants based on brain activity, thereby showing the non-arbitrariness of the annotations. Furthermore, leveraging the structure-based segmentation, we further explore how emotional trajectories relate to formal sections of the concerto and extracted acoustic descriptors, and differences across performances. This work establishes a multi-modal framework for linking annotated emotions, acoustic features, and musical structure, offering a foundation for both musicological analysis and computational modeling of affect in classical music.

References

[1] James Russell. A circumplex model of affect. Journal of Personality and Social Psychology, 39:1161–1178, 12 1980.

[2] Marcel Zentner, Didier Grandjean, and Klaus Scherer. Emotions evoked by the sound of music: Characterization, classification, and measurement. Emotion, 8:494–521, 08 2008

[3] Simon Schaerlaeken, Donald Glowinski, and Didier Grandjean. Linking musical metaphors and emotions evoked by the sound of classical music. Psychology of Music, 50:030573562199123, 03 2021.

[4] Wiebke Trost, Thomas Ethofer, Marcel Zentner, and Patrik Vuilleumier. Mapping aesthetic musical emotions in the brain. Cerebral cortex (New York, N.Y. : 1991), 22, 12 2011.

Preserving Present-Day Basso Continuo Performance through Computational Musicology

Adam Štefunko (Charles University); Jan Hajič, jr. (Charles University)

Improvised harmonic accompaniment is essential to many genres of Western musical heritage, ranging from jazz and folk music to church hymns and even renaissance contrapunto alla mente. How does one conserve and study something that is by design so ephemeral? The answer is, unsurprisingly, in the digital domain. However, the challenge becomes more ethnographic than philological. As professional harpsichordists, we focus on basso continuo, the baroque accompaniment genre of improvising chordal harmonies and polyphonic textures above a notated bass line. This style originates in the early 17th century, but is practiced today within the Historically Informed Performance movement. The rules and guidelines for basso continuo improvisation are found in historical treatises (Viadana, 1602; Muffat, 1699; Mattheson, 1731; Geminiani, 1756), and modern secondary sources (Christensen, 2002). Computational musicology previously focused mostly on studying these rules, including rule-based automatic generation of accompaniments (Wead and Knopke, 2007; Niitsuma et al., 2011), leaving contemporary performance unstudied. We report on the ongoing efforts to document and understand continuo performance. First, we collected ACoRD, the first MIDI dataset of continuo realizations (Štefunko et al., 2025). We have found that adapting MIR methods to a "weakly notated" tradition presents challenges: even aligning the performance to the bass line required moving to a many-to-many alignment paradigm. Alignment enables comparing performances, through the "pivot" of the notated bass line. Working with both historically inspired "griff" representations as well as graph neural networks (Karystinaios and Widmer, 2024), we show that continuo performers today have their distinct styles, and lay the groundwork for computational models of continuo performance. While the HIP movement de-emphasised "authenticity", such models would enable asking the attractive question: are we today actually following the examples of our historical predecessors? But a comparison to the inherently limited historical written sources cannot achieve the broader objective of making sure the skill of continuo is not lost again. Documenting and understanding the present-day practice with computational models might be.

References

[1] Jesper Bøje Christensen. 2002. 18th Century Continuo Playing: A Historical Guide to the Basics. Bärenreiter, Kassel.

[2] Francesco Geminiani. 1756. The art of accompaniament or A new and well digested method to learn to perform the thorough bass on the harpsichord with Propriety and Elegance. Johnson.

[3] Emmanouil Karystinaios and Gerhard Widmer. 2024. Graphmuse: A library for symbolic music graph processing. In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).

[4] Johann Mattheson. 1731. Grosse General-Bass-Schule. Georg Olms Verlag.

[5] Georg Muffat. 1699. Regulae concentuum partiturae. Wien: Musikarchiv Minoritenkonvent (Zentralbibliothek der Österreichischen Minoritenprovinz), MS 1 B 7.

[6] Masahiro Niitsuma, Masaki Matsubara, Masaki Oono, and Hiroaki Saito. 2011. Development of a method for automatic basso continuo playing. Information Processing & Management, 47(3):440–451.

[7] Lodovico da Viadana. 1602. Cento concerti ecclesiastici. Venezia: Giacomo Vincenti.

[8] Adam Wead and Ian Knopke. 2007. A computer-based implementation of basso continuo rules for figured bass realizations. In ICMC.
[9] Adam Štefunko, Suhit Chiruthapudi, Jan Jr Hajič, and Carlos Eduardo Cancino-Chacón. 2025. Basso continuo goes digital: Collecting and aligning a symbolic dataset of continuo performance.

Is Plosive to Nasal Ratio in Poetry Correlated With Affect? The Tasso in Music Corpus Problematizes Claims of Poetic Universals

Joshua Albrecht (The University of Iowa)

Regardless of the language, Auracher (et al.) found that positive and high arousal affect in poetry is correlated with high plosive:nasal ratio in the syllables, whereas negative and low arousal affect is correlated with low plosive:nasal ratio. This correlation was consistent across a surprisingly broad array of texts, ranging from ancient Chinese to modern Russian poetry, suggesting a universal relationship between poetic sound and affect. The current study tested whether traditional markers of musical valence and arousal correlate with plosive:nasal ratios in musical settings of the poems of Italian poet Tasso, one of the most musically influential poets of the late Renaissance, with over 650 musical settings of his poems from over 200 composers. Specifically, this study tested the hypotheses that plosive:nasal ratios in poetic texts positively correlate with melodic interval size, average pitch height, proportion of syllabic:melismatic settings, rhythmic density, rhythmic complexity (nPVI), and proportion of major third vs. minor third sonorities (including incomplete triads). Poems and musical settings were taken from the "Tasso in Music" project (Ricciardi & Sapp). Italian syllables were extracted for each poem, assigned as "plosive," "nasal," or "other," and plosive:nasal ratios were calculated for each poem. Musical features were extracted from the downloaded kern files and calculated using the Humdrum toolkit. Subjective judgments of arousal and valence were also collected for English translations of the poetic texts without musical settings. Correlations were calculated between plosive:nasal ratios, subjective arousal/valence ratings, and the extracted musical features. Plosive:nasal ratios for Italian texts ranged from .6 to 4.6 (mean = 2.0, sd = .76). Most ratios were positive, only 13 using more nasals than plosives. Contrary to the hypotheses, no musical features were significantly correlated with plosive:nasal ratios; all correlations were near 0, consistent with no relationship between ratios and traditional markers of musical affect. Even subjective arousal/valence judgments were uncorrelated with plosive:nasal ratios. These negative results problematize the claim that poetic sounds are universally expressive of affect or translatable into musical expression. The results would also be consistent with musicians not explicitly encoding lyrics' valence into musical settings, suggested by Palomeque & de Lucio (2021), though that idea would contradict centuries of music scholarship.

Spectral Influence on Emotional Affect in Music Listening: An Empirical Study of Frequency Balance and Perceived Mood

Sravan Palla (ICFAI); Kamalakar Karlapalem (IIIT HYDERABAD)

A preliminary empirical study examining whether or not the spectral frequency balance of music — the distribution of energy across the frequency spectrum — can independently influence emotional affect in listeners. Six original instrumental tracks were processed into three versions — Sad, Neutral, and Happy — using parametric equalisation, with all melodic, rhythmic, and harmonic content constant. Participants listened to all three versions in a randomised order and rated their emotional response after each. A mean audibility rating of 9.0/10 confirmed that spectral differences were clearly perceivable, and 100% of the people surveyed identified the Happy version as clearly brighter. However, responses on emotional belief and willingness to reuse the tool reveal significant ambivalence: while instantaneous spectral perception is demonstrably affected, participants remained largely unconvinced that EQ manipulation could or should change the emotional core of a piece of music. These findings raise important questions about the limits of technical intervention in art reception, and engage with the artistic principle — articulated by the composer Hans Zimmer — that music should invite the listener to feel something rather than instruct them what to feel.

A Schema-Aligned Ontology and Automatic Lifting Methodology for Representing MEI as Granular Linked Data: Benchmarking SPARQL Responsiveness in Real-Time MIR Tasks

Joshua Stutter (Newcastle University)

The Music Encoding Initiative (MEI) format (2026) is an XML hierarchical music encoding representation that enjoys wide support by musicologists and MIR researchers due to its commitment to open standards, ease of encoding, and semantic emphasis (Library of Congress, 2024). Real-time retrieval and corpus analysis directed by a user query is possible using MEI by loading the corpus' XML into an analysis package such as music21 and iterating at query time (Cuthbert et al., 2011). For example, "find all melodic intervals C5—E5" requires traversing every note, parsing its pitch, and filtering notes with C5 followed by E5. This scales linearly: a sizeable corpus makes queries slow. Ratta & Daga (2022) have shown that this bottleneck can be sidestepped by exposing symbolic music as RDF using SPARQL-Anything (Asprino et al., 2023) and querying it declaratively, validating their MusicXML approach against music21 for correctness. Their work, however, established only feasibility: querying the raw façade output of SPARQL-Anything rather than data lifted into a semantically enriched graph, targeted MusicXML rather than MEI, and did not measure query speed, which is the property that motivates a database approach over linear iteration. This paper builds directly on Ratta & Daga, addressing each limitation. MEI is automatically converted into an equivalent, granular linked-data ontology: an RDF lifting query and SHACL validation shapes are generated from the MEI schema, enriching the basic façade triples, and loaded into Apache Jena as a natively indexed graph. Storing the enriched data as a graph in a triplestore allows fast lookup of arbitrary, cross-cutting queries, with sub-linear growth. Benchmarking against music21 demonstrates significantly improved throughput for several real-world MIR tasks, including interval and n-gram retrieval, confirming the speed gains that have been heretofore unmeasured. Using MEI as linked data is therefore recommended for large music corpora at scale.

References

[1] Asprino, L., Daga, E., Gangemi, A. and Mulholland, P. (2023). Knowledge Graph Construction with a Façade: A Unified Method to Access Heterogeneous Data Sources on the Web. ACM Transactions on Internet Technology, 23(1). https://doi.org/10.1145/3555312.

[2] Cuthbert, M.S., Ariza, C. and Friedland, L. (2011). Feature Extraction and Machine Learning on Symbolic Music Using the music21 Toolkit. In: Proceedings of the 12th International Society for Music Information Retrieval Conference. [online] pp.387–392. Available at: https://ismir2011.ismir.net/papers/PS3-6.pdf.

[3] Library of Congress (2024). Music Encoding Initiative (MEI) format family. [online] Sustainability of Digital Formats. Available at: https://www.loc.gov/preservation/digital/formats/fdd/fdd000502.shtml.

[4] Music Encoding Initiative (2026). MEI: The Music Encoding Initiative Guidelines. [online] Zenodo. https://doi.org/10.5281/zenodo.14795519.

[5] Ratta, M. and Daga, E. (2022). Knowledge Graph Construction from MusicXML: An Empirical Investigation with SPARQL Anything. In: Music Heritage Knowledge Graphs (MHKG): 21st International Semantic Web Conference. [online] Available at: https://oro.open.ac.uk/85326/1/Music_Knowledge_Graphs_Paper%20(3).pdf.

Choral meets Machine Learning: AI Gregorian Chant with the Regensburger Domspatzen

Andreas Haberl (Regensburger Domspatzen); Benedikt Dolderer (Regensburger Domspatzen)

With a history spanning over 1,050 years (founded in 975), the Regensburger Domspatzen are the oldest boys' choir in the world; starting in fifth grade, we regularly sing Gregorian chant in Regensburg Cathedral. Gregorian chant is the oldest musical form of the Western world—monophonic, without a fixed meter, following the flow of the Latin text, and based on modal church modes—and at the same time a unique, yet largely unseen, cultural treasure. Based on this lived practice, we have developed an algorithm that learns the melodic grammar of the chant from approximately 9,000 historical training data points and uses it to generate new, stylistically coherent melodies. The chorales are tokenized in the GABC format so that the modal musical language can be modeled sequentially; on this basis, we compare two neural sequence models: a stacked LSTM and a Transformer. Since a purely data-driven model only partially captures the formal rules of the chorale, a specially developed, music-theoretically grounded post-processing step supplements the model's outputs. It controls the degree of conventionality and melismatic ornamentation and ensures a modeappropriate conclusion on the finalis. Beyond mere generation, this approach opens up new perspectives in musicology. For example, the model reconstructs the missing ending of a torn parchment fragment (circa 1300), allowing a melody to be sung again after some 700 years. Our goal is to show how AI can preserve and creatively develop culture, rather than replace it! Audio examples from a Schola der Domspatzen choir can accompany this presentation.

Poster Session D

Vibrato as Rhythmic Subdivision: Cross-Genre Evidence for Vibrato-Tempo Coupling

Michael Van Bebber (Independent Researcher)

This study presents cross-genre evidence for a previously undescribed relationship between vibrato rate and musical tempo. While vibrato rate is conventionally reported in Hz as a fixed individual characteristic (typically 5–8 Hz, equivalent to 300–480 BPM), the present research demonstrates that vibrato functions as a rhythmic subdivision of the musical pulse, with performers across genres, instruments, and historical periods synchronizing vibrato oscillations to the metric grid. Because vibrato is shown here to be a rhythmic rather than purely acoustic phenomenon, rates are reported throughout in BPM — the unit in which the musical pulse itself is measured. A corpus of hundreds of commercial recordings spanning jazz, soul, rock, country, and classical — representing artists across more than a century of recorded music — was analyzed for song tempo and vibrato rhythm. Vibrato subdivisions were identified aurally by a doctoral-trained jazz performer and educator with extensive experience in rhythm and pulse perception, supplemented by spectral analysis. Six discrete tempo zones emerged from aggregate analysis, each corresponding to a specific subdivision of the beat: triplet sixteenth notes at approximately 50–75 BPM, sixteenth notes at 75–100 BPM, triplet eighth notes at 100–150 BPM, eighth notes at 150–200 BPM, triplet quarter notes at 200–300 BPM, and quarter notes at 300–400 BPM. These boundaries represent aggregate averages rather than fixed thresholds — individual performers and genres show variation around these norms. Alternate zone boundaries follow doubling relationships (e.g., 50–75 and 100–150 BPM), reflecting the octave-like relationship between same-subdivision-type zones. The subdivision mechanism maintains absolute vibrato rate within approximately 300–450 BPM regardless of song tempo — the empirical signature of a comfortable motor-oscillation range preserved by shifting subdivisions as tempo changes. The zone structure's consistency across genres, instruments, and eras suggests vibrato-tempo coupling reflects a fundamental neurocognitive mechanism of metric entrainment rather than a learned stylistic convention. These findings extend the inconclusive results of Desain et al. (1999), whose narrow experimental tempo range fell within a single zone. Preliminary analysis indicates filtering the corpus by artist or genre yields tighter tempo correlations, suggesting aggregate zones represent a superimposition of stylistically coherent sub-patterns. The framework has been applied pedagogically in a vibrato training application currently in use.

References

[1] Desain, P., Honing, H., Aarts, R., & Timmers, R. (1999). Rhythmic aspects of vibrato. In P. Desain & W. L. Windsor (Eds.), Rhythm Perception and Production (pp. 203–216). Lisse: Swets & Zeitlinger.

The Interactive Drunken Sailor: Exploring Interactive Visualisations of Variants in Digital Music Editions

Anna Plaksin (Paderborn University); Tobias Bachmann (Paderborn University); Adrian Nachtwey (Paderborn University)

Studying music across time, space, and cultures involves the investigation of change — the change of music cultures. One main witness that allows the observation of change in music cultures lies in the music itself. The act of passing down changes music, whether disseminated orally or in writing. Reconstructing dissemination paths enables us to study music as one of many subjects of cultural change and provide deeper insights into its role in cultural and political discourse — not just by tracing agents of dissemination, but also by following the music itself. Since human reproductions of text – literature or music – inevitably introduce accidental or intentional changes, studying this variation is the traditional method for reconstructing text and music traditions. Dealing with textual traditions and their variation is an integral component of music philology and editorial practice. Editions frequently conceal variation by privileging a preferred text — thus rendering it inaccessible as an object of study. While digital representations of music editions have targeted the visualisation of divergent versions for almost two decades, reusable solutions for comparing, describing, and visualising variants across more than two equivalent versions are still absent. Making variation investigable requires new philological approaches in both editorial practice and digital presentation. We present the Interactive Drunken Sailor (IDS), an application that enables users to create their own version of the well-known sea shanty 'What shall we do with a Drunken Sailor?' by selecting various readings from four different transmissions. The IDS is a conceptual piece to explore and showcase dynamic music scores. Based on the IDS, we discuss three conceptual questions: 1. How can digital music editions be created without establishing a primary text? 2. How can musical variants be visualised and made explorable? 3. How can the interdependence of musical variants — where a single notational sign may affect the entire text — be addressed?

Time to Align! Workflows, Applications, and Exchange Format for a Model of Musical Timelines

Davide Andrea Mauro (Paderborn University); Axel Berndt (Paderborn University); Carlos Eduardo Cancino-Chacón (Johannes Kepler University Linz); Simon Dixon (Queen Mary University of London); Anne Foo (Independent Researcher); Mark Gotham (King's College London); Johannes Hentschel (Anton Bruckner University); Patricia Hu (Johannes Kepler University Linz); Maik Köster (Anton Bruckner University); Felipe D. Martins (Anton Bruckner University); Meinard Müller (International Audio Laboratories Erlangen); Markus Neuwirth (Anton Bruckner University); Alexander Pacha (TU Wien); Kevin R. Page (University of Oxford); Silvan Peter (Johannes Kepler University Linz); Egor Polyakov (HfM Weimar); Laurent Pugin (RISM); David Weigl (University of Music and Performing Arts Vienna); Christof Weiß (University of Würzburg); Gerhard Widmer (Johannes Kepler University Linz)

Aligning corresponding music representations, such as graphical scores (images, scans), logical encodings (MIDI, MEI), and physical recordings (audio, sensor data), is central to Music Information Retrieval (MIR) and Digital Musicology. However, each representation imposes its own native coordinate system for musical time: Integrating them requires costly alignment efforts, and the resulting data is typically stored in heterogeneous formats that are hard to reuse. To address this, we recently introduced Time to Align! [1], a cross-domain model and Python library aimed at facilitating the creation, storage, reuse, and conversion of alignments between musical timelines. Focusing on practical applications and interoperability, we detail a dataset import pipeline designed to reconcile disparate coordinate systems, such as continuous time in seconds, discrete ticks in MIDI, and identifiers in MEI. We demonstrate how diverse formats are ingested and translated into a unified ecosystem of typed timelines, connected by explicit conversion maps and match claims. Finally, we showcase the flexibility of this framework through a series of example applications. Our main contribution consists of a first draft of a data exchange format designed to round up our model implementation and eliminate isolated data silos. Based on the Parquet standard, this format standardises how alignments are structured, promoting them to first-class data entities and enabling researchers to publish, share, and reuse alignment data independently of the underlying musical datasets it connects. We invite attendees to explore our documentation, tutorials, source code, and runnable notebooks via https://timetoalign.github.io/ and encourage community feedback and discussions for the ongoing development of the initiative.

References

[1] Johannes Hentschel, Axel Berndt, Carlos Cancino-Chacón, Simon Dixon, Ann Foo, Mark Gotham, Patricia Hu, Maik Köster, Felipe D. Martins, Davide Andrea Mauro, Meinard Müller, Markus Neuwirth, Alexander Pacha, Kevin R. Page, Silvan Peter, Egor Polyakov, Laurent Pugin, David Weigl, Christof Weiß, and Gerhard Widmer. "Time to align! modelling musical timelines for music information retrieval and digital musicology." Accepted in: Transactions of the International Society for Music Information Retrieval (TISMIR), 2026.

Symbolic Music Calculation Beyond 12-ET With xenharmlib

Fabian Vallon (Independent Researcher)

Xenharmlib is a free / open source python library that makes it possible to analyze musical objects like pitches, scales and interval sequences in all equal division tunings. These include the contemporary Western 12-ET, the modern Arabic 24-ET, the Turkish 53-ET as well as novel tunings like Bohlen-Pierce that divides a tritave into 13 equal steps. Divisions and equivalency intervals can be freely chosen, so even tunings like 31-ET (used e.g. in the work of Renaissance composer Nicolò Vicentino) or 72-ET are possible. Xenharmlib implements Kite Giedraitis' "Ups and Downs Notation", a superset of Western notation that adds step accidentals to the Western sharp and flat accidental system. On top of that it provides basic analytical utilities like searching possible chords in scales, pitch class analysis, posttonal prime form calculation or finding connecting triads between different scales. The presentation aims not only to introduce the current features of xenharmlib, but to find collaborators for its ongoing development which is currently focused on support for music in Just Intonation (like many non-western music traditions or the music of Harry Partch). The library is released under the GNU Public Licence v3 and documentation can be found under the address `https://xenharmlib.readthedocs.io/en/latest/`.

Tuning Into Generation Z: A Value-Sensitive Design Approach to Redesigning Muziekweb, a Digital Music Archive, Supporting Engagement with Music Across Time, Space, and Cultures

Shivani Aggarwal (Utrecht University); Karlijn Dinnissen (Utrecht University); Hanna Hauptmann (Utrecht University); Anja Volk (Utrecht University)

Cultural heritage platforms face growing pressure to remain relevant to younger audiences with increasingly high expectations for online experiences (Konstantakis and Caridakis 2020). Studies on such young audiences largely focus on either museum-based experiences in broad cultural heritage contexts or media consumption habits in general, while comparatively little attention has been given to music cultural heritage or what especially highly engaged listeners value in platforms dedicated to the exploration of music culture, beyond what mainstream streaming services offer. Muziekweb, one of the Netherlands' largest digital music archives and part of the Dutch Institute of Sound and Vision, is preparing a major redesign aimed at reflecting contemporary music culture and attracting new users, particularly Generation Z (born 1997–2012) (Wandhe 2024). Applying a Value Sensitive Design approach (Friedman et al. 2013) to bridge this gap, semi-structured interviews were conducted with Generation Z music enthusiasts and Muziekweb professionals to elicit stakeholder values around music, cultural heritage, and digital platforms. Findings reveal that both groups share a resistance to algorithmically and commercially driven music experiences and a preference for human connection, contextual depth, and distinctive content. Generation Z interviewees specifically express frustration with repetitive offerings of streaming platforms, instead seeking to discover unfamiliar music within its broader cultural, historical, and social context, deepening the meaning of their music experiences. The identified values were translated into design requirements and realized through an interface, which may be implemented by any platform seeking to better align with younger audiences' values. Moreover, we report our findings on Generation Z's interest in learning about music across time, space, and culture. They show that computational musicology may support access to the content and context of digital musical collections, enriching educational experiences for younger audiences through approaches informed by musicological knowledge, including digital storytelling and immersive technologies.

References

[1] Friedman, Batya, Peter H. Kahn Jr., Alan Borning, and Alina Huldtgren. 2013. "Value Sensitive Design and Information Systems." In Early Engagement and New Technologies: Opening Up the Laboratory, 55–95. Springer.

[2] Konstantakis, Markos, and George Caridakis. 2020. "Adding Culture to UX: UX Research Methodologies and Applications in Cultural Heritage." Journal on Computing and Cultural Heritage 13 (1): 1–17.

[3] Wandhe, Priyanka. 2024. "The New Generation: Understanding Millennials and Gen Z." SSRN Electronic Journal. `https://ssrn.com/abstract=4716585`.

Task-Oriented Evaluation for Automated Music Simplification

Andrew McLeod (Fraunhofer IDMT); Hanna Lukashevich (Fraunhofer IDMT); Jakob Abeßer (University of Bamberg)

The task of music simplification involves the conversion of a piece of music into a version that is, in some sense, easier or simpler [1]. While some models have been proposed for the task, no standardized evaluation metric or procedure has yet been established. We argue that the reason for this is the wide range of use cases for which music simplification could be applied. A major class of use cases for music simplification is that of making music easier to play, with different goals: allowing a musician with a lower skill level to play something that still captures the essence of the full piece; helping a musician to learn the full version of a piece by learning progressively more difficult versions [2]; or improving music accessibility by, e.g., reducing hand span for piano or limiting the fret range of a piece for guitar. A second class involves the abstraction of a piece of music for the purpose of a deeper understanding or analysis (e.g., [3]). Such a simplification might result in a version which could help musicologists in producing a harmonic analysis of a given piece, or a version which allows performers and accompanists to better improvise along with it. Finally, a third class considers the use of music simplification for the purpose of some downstream MIR task. Examples include query-by-humming [4], where a user might query using a simplified version of the desired melody; or music generation, where (pre-)training a model on abstracted versions might help to avoid picking up genre-specific biases from the training data. Each of these use cases and goals is distinct, and music simplification models designed for one might perform poorly on another. We argue therefore, that use case and goal specific evaluation procedures are required. We consider how this might look for each with the help of initial results from a music simplification model.

References

[1] P. Ramoneda, M. Suzuki, A. Maezawa, and X. Serra, "Difficulty-aware score generation for piano sight-reading," Expert Systems with Applications, vol. 321, p. 132088, 2026, issn: 0957-4174. doi: https://doi.org/10.1016/j.eswa.2026.132088.

[2] P. Ramoneda, E. Parada-Cabaleiro, D. Jeong, and X. Serra, Difficulty-controlled simplification of piano scores with synthetic data for inclusive music education, 2025. arXiv: 2511.16228.

[3] P. Ericson, Y. Rammos, and M. Rohrmeier, "Musereduce: A generic framework for hierarchical music analysis," in Music Encoding Conference 2022, Halifax, Canada, May 19-22, 2022, Music Encoding Initiative, 2023, pp. 40–51.

[4] N. Kosugi, Y. Nishihara, T. Sakata, M. Yamamuro, and K. Kushima, "A practical query-by-humming system for a large music database," in Proceedings of the eighth ACM international conference on Multimedia, 2000, pp. 333–342.

Rhyme, More Pervasively Than Harmony, Drives Perception of Closure in Sung Poems: An Exploration of the Interaction Between Phonological and Tonal Hierarchy in Sung Music

Juan Felipe Pérez Ariza (University of Vienna); Tudor Popescu (University of Vienna)

Music and Language have been compared as structurally analogous cognitive systems, but the appropriate level at which this comparison should be drawn remains contested. Phrasal syntax in language, and the structural organisation of Western tonal music, have been jointly studied, leading to important insights about perception of well-formedness and structure across both domains. Here, we propose to shift the focus of language from phrasal syntax to phonology to draw the comparison with musical tonality. We argue that using rhyme within language, understood both as a hallmark of closure in poetry, and as a hierarchical relationship between phonological elements (namely, syllables), might offer a better level of comparison with tonal hierarchy than phrasal syntax. Using two-line, sung poems we explored the perception of closure in language and music by independently manipulating rhyme in lyrics, and harmonic cadence in music. Using a two-by-two design, we exposed 30 individuals to congruent and incongruent stimuli, and we probed their perception of closure in each domain, resulting in four closure possibilities: no closure, only language closure, only music closure, or full closure. Preliminary results suggest that participants identified full closure when stimuli contained rhyme and cadence, but, interestingly, also tended to perceive a full closure when there was rhyme but no cadence in the stimulus. These observations suggest that the perception of rhyme drives the perception of closure in music much more pervasively than the opposite, in a potential asymmetrical cross-domain influence. This asymmetry supports our proposal that phonological structure might offer a more productive level of comparison with tonal hierarchy than phrasal syntax in Western listeners, as it suggests both systems interact asymmetrically at the level of closure perception rather than sharing equivalent structural weight. This, however, also raises the question with regards to the perception of closure in non-Western traditions.

Towards a Corpus of Kurdish Folk Music

Ameer Majeed (Charles University); Jan Hajič, jr. (Charles University)

The Kurds are one of the largest ethnic groups without an independent nation-state, predominately located in Türkiye, Iran, Iraq, and Syria. Throughout decades of systematic assimilation and censorship of the Kurds within these nation-states, linguistic, cultural, and musical expressions were suppressed [6, 4]. In contrast to other musical traditions, Kurdish folk music lacks documentation and research. While some ethnographic collections exist [2, 3, 1, 5], they are far from comprehensive. In this ongoing project, we have carried out initial steps towards collecting a digital corpus of Kurdish folk music and its typology. As few melodies are documented, we focus on lyrics; specifically in Kurmanji Kurdish, for which most extensive resources exist. We transcribe the lyrics of 421 songs from the two largest ethnographic collections: Kilamê Cimeta Kurda [3] and Folklora Kurdî [2]. Folklora Kurdî proposes a typology of Kurdish folksongs; however, our survey revealed it has limited relevance to contemporary listeners, including those familiar with the songs. Furthermore, few songs from the collections are known today, even to experts and musicians. As a second layer representing contemporary Kurdish music, we collect another 3842 song lyrics from the community-based website LyricsTranslate. However, here we must re-examine the term "folk" music in light of contemporary musical practices. We start by cross-referencing lyrics from the early and contemporary sub-corpora, and identify some textual elements that may serve as stable markers of music that is constitutive to Kurdish identity. The project has so far been mostly re-discovering issues of documenting a marginalised tradition for which no established ethnomusicological framework exists, not even a basic song typology — but what is novel is the opportunity to do so with the affordances of scale and multi-modality of the digital domain and computational methods.

References

[1] N. Christensen and D. Christensen. Kurdish Folk Music from Western Iran, 1966.
[2] Cigerxwîn. Folklora Kurdi: Kurmancî. Roja Nû, Stockholm, 1988.
[3] Heciyê Cindî. Kilamê Cimeta Kurda. Lîmûș, Yerevan, 2008.
[4] David Mcdowall. A Modern History of the Kurds. Tauris, London, 2004.
[5] Ralph Solecki. Kurdish Folk Songs and Dances, 1955.
[6] Welat Zeydanlioglu. 'The White Turkish Man's Burden' : Orientalism, Kemalism and the Kurds in Turkey. Cambridge Scholars Publishing, Newcastle, 2008

The Evolution of Hip Hop: Variation in Flow Triggered by Environmental Change

Emery Thul (Princeton University); Asif Ghazanfar (Princeton University)

Since its origins at house parties in the Bronx, New York during the late 1970s, hip hop has changed dramatically across time and space. Technological, economic, regional, and other environmental forces drive these changes, contributing both to hip hop's global significance and local idiosyncrasies via cultural evolutionary processes. To quantify these processes, previous corpus studies measured changes in the complexity and variation in "flow"—the rhythmic delivery of rap verses relative to the beat (Ohriner, 2016, 2019; Condit-Schultz 2016; Waller 2016; Duinker 2020). While these approaches generated robust descriptions of how hip hop has changed over time, they often overlook the environmental factors that explain these changes. In this paper, we address not only how flow changes over time, but how specific events (e.g., the advent of the digital sampler) or geographic regions (e.g., East versus West Coast) influenced these changes. We constructed a corpus containing 400 hip hop tracks from 1980 to 2020 from four major hip hop regions in the United States (East, West, South, and Midwest). From these tracks, we extracted vocal and drum stems before computing quantized inter-onset intervals (IOIs). Using these IOIs, we generated phase portraits to qualitatively describe the evolution of metrical complexity in flow over time and across regions. Additionally, we explored the degree to which the onsets of the vocal flow align with the underlying meter and drum tracks to quantitatively measure the complexity of flow. Our preliminary results suggest that these methods capture significant temporal and regional variation that can be traced to the influence of specific environmental pressures.

Reconstructing the Historical Ontology of a Popular Music World: A Multi-Agent LLM Pipeline Applied to an Oral History Archive from Postwar Taiwan

Ming Cheng (Academia Sinica); Yun-En Tsai (Academia Sinica); Li Su (Academia Sinica)

Popular music history has long been governed by universalising narratives that obscure music's dimensions as labour, collaboration, and technical mediation. This study addresses one such occluded axis: in postwar Taiwan's popular music industry, the backstage practitioners of musicking as a technical practice have long remained narratively peripheral. This peripherality is produced by familiar narrative structures of power, not by any deficit on the part of front-stage figures. From an oral history archive of roughly three hundred volumes, a set of collaborating on-device large language model (LLM) agents extracts directed relational claims of mentorship, introduction, employment, collaboration, and conflict, each carrying a temporal anchor and an epistemic-status marker that records how firmly the testimony supports it. From these claims we assemble a historical knowledge ontology as a directed graph whose nodes span front-stage and backstage practitioners. On this ontology we pursue two analytic moves. First, by encoding directionality, we surface role asymmetries that undirected structures obscure. Second, by timestamping the directed network, we trace the temporal dynamics of group closure and reconfiguration. To this end, we apply direction-sensitive measures including PageRank, HITS hub/authority scores, and reciprocity. The historiography of popular music in Taiwan presents a methodological gap: individual archives, memoirs, and interviews accumulate without an established path toward a structurally analysable ontology. Our approach builds the ontology bottom-up from the practitioners' own accounts, with its shape determined by collaborative networks that actually existed rather than by the assumptions of established industrial or media narratives. Language models, together with the extraction design, serve as mediators: prompted by scholarly questions, they traverse the ontology to generate situated historical narratives, while the interpretive responsibility remains with human scholars. The study thereby points toward a convergence of computational method and historical musicology. Its final product is a historical ontology open to continual re-entry.

Uncovering the Evolution of Music: A Segmentation and Representation Learning Framework for Cross-Composer Analysis of Musical Structure

Benjamin Oh (Lynbrook High School); Benjamin Zhang (Lynbrook High School)

Musical structure has historically evolved through the transmission and adaptation of ideas across generations of composers. However, analyzing such processes still remains a central challenge in computational music analysis, as existing approaches often only emphasize harmonic progression and chord-level representations, and are limited in their ability to capture large-scale structural and motivic development. We present an unsupervised framework for studying musical structure across time periods through graph-based segmentation and representation learning. MIDI and MusicXML scores are converted into temporal note graphs, segmented using graph-based changepoint detection inspired by G-PELT, and refined into phrase-like units. These segments are then embedded using a contrastive representation model designed to preserve motivic identity under variation while reducing dependence on orchestration-heavy surface features. The resulting embeddings support analyses of motivic recurrence, transformation, and stylistic similarity both within and across composers. Using a dataset of late-Romantic to contemporary composers such as Brahms, Wagner, and Schoenberg to revisit the historical claims made in Schoenberg's Brahms the Progressive, this work demonstrates that, across motif representations derived from phrases, Schoenberg is consistently closer to Brahms than to Wagner, with a cosine similarity of 0.9523 to Brahms compared to 0.9337 to Wagner. This result holds across multiple metrics, including motif-cluster distributions, nearest-neighbor retrieval, and recurrence structure, suggesting that the computational analyses of such metrics can recover historically meaningful relationships that align with long-standing musicological theories of inheritance. More broadly, this model provides a reproducible approach for studying musical influence and stylistic continuity across historical periods. By focusing on how musical ideas are transformed, repeated, and organized, our work offers new tools for investigating the evolution of musical structure across time and lays the groundwork for future studies involving larger and more culturally diverse musical corpora.

Effects of Metrical Granularity on Adaptation to Tempo Change

Joris Monnet (EPFL); Xiaoxuan Wang (EPFL); Martin Rohrmeier (EPFL)

Beat tracking is the ability to perceive a periodic pulse in a rhythm and synchronize with it. Previous research has linked beat tracking to several internal mechanisms: error correction, multi-level oscillator entrainment, multiple temporal references within a metrical hierarchy, and continuous Bayesian phase-and-tempo inference. Yet whether the metrical hierarchy shapes adaptation during tempo updates, when predictions need to be revised, remains untested. The subdivision density sets the number of internal timepoints per cycle. When the tempo changes, subdivision density may act in two antagonistic ways: (1) it sharpens the new-tempo estimate by providing higher resolution; (2) it strengthens the prior on the old tempo, slowing the update. We ask which effect dominates and whether a densely structured grid that is only partially presented can still be fully internalized, yielding adaptation comparable to that observed under full presentation. We will use a sensorimotor synchronization paradigm whose test-phase surface events are matched across conditions, so any adaptation difference must arise from internal representation. A priming phase induces a metrical grid through an accented high-pitched downbeat and selectively sounded weak lower-pitched beats; the test phase leaves only the downbeat. A gradual three-cycle tempo ramp between 80 and 100 BPM, direction concealed, occurs at a fixed cycle boundary while participants tap the beat. Granularity spans a four-pulse grid, a three-pulse grid, a two-pulse grid, and no subdivision; the grids are fully sounded or have some of their weak beat missing, testing whether subdivision location matters. Adaptation will be quantified by transition-tracking RMSE and overshoot. If the denser grid speeds adaptation, Bayesian evidence accumulation dominates; if it slows it, the prior wins. If partial presentation matches full presentation, listeners reconstruct the grid from sparse cues. Together they would offer the first joint test of how granularity supports dynamic synchronization and how it is internalized.

Symbolic Transformer Models for Distinguishing Between Earlier and Later Styles of Solo Piano Music of Three Romantic Composers

Ruach Jacobsohn (University of Pretoria); J. Pieter Jacobs (University of Pretoria)

The closely related tasks of style-based composer classification and authorship attribution from symbolic music scores have received increased attention in recent years (Simonetta, 2025). Composer identification is more reflective of composers' individual styles if the composers belong to the same period; authorship attribution is best carried out when based on score-based symbolic data, (as opposed to performance-based data). The present study focuses on a related, more fine-grained (and, as far as we could establish, fairly unexplored) problem, namely that of distinguishing between the phases of the style of an individual composer. For example, compared to his earlier works, the late piano works of Johannes Brahms exhibit fluidity of structure, and an obfuscation of the distinction between melody and harmonic support (Bozarth and Frisch, 2001). Regarding Franz Liszt, Charles Rosen (Rosen, 1999) observes that, in contrast to his early works, many of his late piano works presage Debussy and early-twentieth century atonality. A dataset of solo piano scores (MusicXML) was curated from Musescore and converted into an adapted OctupleMIDI token representation (Zeng et al., 2021), capturing the time signature, tempo markings, bar, position within bar, pitch, note length and bar length for all music events. The beat-per-minute values were not reliably available across all scores, therefore, to improve consistency, tempo markings (Adagio, Grave, etc.) were implemented instead. Additionally, a bar length feature was added, as the time signature feature does not reliably reveal the objective bar length across all instances in a score. Earlier-style/later-style classification was performed using a (non-pre-trained) encoder transformer, implementing binary classifiers for composers individually with segmented input data. Model performance was evaluated using k-fold cross-validation. Experimental results suggest that the approach is promising, with balanced accuracies achieved of up to 83% (Brahms). These results indicate that the transformer architecture is capable of learning sophisticated intra-composer stylistic differences.

References

[1] Bozarth, G.S. and Frisch, W. (2001). Brahms, Johannes. In D. Roote (Ed.), Grove Music Online, Oxford University Press. `https://oxfordmusiconline.com`.

[2] Rosen, C. (1999). The Romantic generation, Fontana Press.

[3] Simonetta, F. (2025). Style-based composer identification and attribution of symbolic music scores: A systematic survey. Transactions of the Society for Music Information retrieval, 8(1), 213–235.

[4] Zeng, M., Tan, X., Wang, R., Ju, Z., Qin, T. and Liu, T.Y., 2021, August. MusicBERT: Symbolic music understanding with large-scale pre-training. In Findings of the Association for Computational Linguistics: The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021) (pp. 791-800).

Perceived Harmonic Tension in Jazz: Tonal Geometry versus Learned Surprisal

Léo Bruneau (EPFL); Loïc Domingos (EPFL); Xiaoxuan Wang (EPFL); Edward Hall (EPFL); Martin Rohrmeier (EPFL)

Harmonic tension is central to theories of musical expectation and emotion, yet accounts of what makes a chord feel tense diverge. Tonal-geometric theories characterize tension by a chord's distance from a stable region within a pitch space; predictive accounts characterize using chord surprisal, given a listener's learned model of harmonic syntax. Current computational models of harmonic tension focus on Western-classical music, yet jazz—with its extended dominants and modal interchange—is a tonal idiom where their predictions may diverge. Here we ask whether perceived tension in jazz is better explained by tonal geometry or by learned surprisal. In an online experiment, continuous tension judgments will be measured from listeners (N=100), each rating 25 jazz progressions (M=34.5 chords, SD=1.7) on a slider. These judgments are compared against two models: a geometric tonal-tension model built from Lerdahl-style tonal-spatial distances, and a learned-surprisal account that estimates each chord's predictability from its harmonic context. We take a large pretrained symbolic-music model (MusicBERT) as a baseline for the learned account, but do not restrict ourselves to it. We will also evaluate causal, jazz-trained sequence models, whose surprisal more directly reflects a listener's forward expectations. Listeners' moment-to-moment rating are aggregated at the chord level to give one shared tension value per chord. Because the models are more likely to agree on diatonic motion, but diverge on idiomatic events, a meaningful comparison can be made between them. We expect tonal-geometry to explain much of the variance on diatonic and cadential events, while learned-surprisal adds variance on jazz-idiomatic chords where classical geometry under-predicts felt tension; this would imply that perceived tension reflects the statistical familiarity with an idiom, as much as tonal-spatial structure. Musical sophistication may modulate this contrast, with idiom-experienced listeners tracking surprisal while others default to tonal-distance cues.

The Interactivity Gap: Deep Representations vs. Acoustic Features in Neurodivergent Musicing

Frederico Pedrosa (Universidade Federal de Minas Gerais); Marina Freire (Universidade Federal de Minas Gerais); Aline André (Universidade Federal de Minas Gerais)

This study explores the computational modeling of musical interactivity, a core construct in music therapy and cognition. While manual microanalysis provides high-fidelity data, its artisanal burden remains unscalable for public health. We propose a proof-of-concept for automating interactivity prediction by comparing two paradigms: deep representation learning (MERT) and hand-crafted acoustic features. Using 240 clinical excerpts (120 minutes), we extracted latent interactivity scores via Confirmatory Factor Analysis of the Nordoff-Robbins Scales (ω=0.876) to establish a psychometrically sound ground truth. We implemented a "Tournament of Models" using Regularized Regression and Non-linear Ensembles. Our results demonstrate a stark representational dissociation: MERT excels in improvisational, dialogic contexts (R2=0.827), tracking high-level musical semantics and interaction shared between patient and music therapist. Conversely, hand-crafted features perform better in task-based neurological frameworks (R2=0.677), where interaction is primarily motor-functional and rhythmic. Notably, we identified a generalization collapse (R2=0.029) and fusion failure when forcing cross-context model transfers. This mathematical lack of convergence indicates that musical interactivity is not a monolithic acoustic signal but a context-dependent cognitive construct. While deep learning captures the dialogic essence of musicing, an audio-only “ceiling” exists. We conclude that audio-only descriptors reach a limit of clinical interpretability when separated from physical gestures. This underscores the need for multimodal architectures, accounting audio and motion, and framework-aware AI to mitigate clinical blind spots in neurodivergent assessment. Finally, we discuss the ethical necessity of privacy-preserving, on-device processing for high-stakes clinical MIR.

Tracing the Evolution of Ragas Over Time Through Phrase Prevalence Analysis

Subashri V (TIFR Centre for Interdisciplinary Sciences, Hyderabad); Thomas Nuttall (Universitat Pompeu Fabra)

I am a postdoctoral fellow in statistical physics, and a Carnatic musician with over a decade of concert performance experience. In collaboration with Thomas Nuttall, I have been developing computational methods to trace the evolution of Carnatic Rāgas. Carnatic music, the classical music of South India, has a rich repertoire of Rāga-based compositions spanning the seventeenth century to the present, transmitted through lineages that emphasize musical fidelity. Computational studies of Carnatic Rāgas have investigated Rāga recognition (1), melodic similarity (2), and characteristic phrase discovery (3); however, quantitative studies of the historical evolution of phrase vocabularies of Rāgas remain limited. We build exhaustive phrase annotations in various Rāgas from an extensive corpus of Carnatic compositions, grouped into three periods: Group 1: 17th to 18th century Group 2: post-1850 up to 1900 Group 3: 1900 to present day Although this classification is not chronologically rigid, it represents a musical chronology, wherein Group 1 is dominated by compositions of the musical Trinity (4), their contemporaries and immediate predecessors, Group 2 largely by disciples of the Trinity, and Group 3 by modern composers. Using the annotations, we investigate changes in the phrase vocabulary of Rāgas over these time periods. Analysis involves computing the relative prevalence of annotated patterns in the three groups using a metric derived from term frequency–inverse document frequency (TF-IDF) weighting to understand how the Rāga has developed over time. We present preliminary findings from this framework and discuss phrase-level differences in multiple Rāgas across time periods, with their possible musicological interpretations, drawing both from my experience as a musician and existing theories. For example, consider Rāga Bhairavi. We find differences in the prevalence of several characteristic phrases across periods. In particular, phrases containing D_2 in ascent, such as D_2NS, show highest prevalence in twentieth-century compositions, supporting modern practice despite earlier theoretical descriptions of Bhairavi as employing only D_1 (5). Conversely, the ascending phrase PD_1ND_1P is most prevalent in nineteenth-century compositions, while the phrase ND_1, P is most prevalent in modern compositions. The phrase ND_1, P is commonly associated with Rāga Manji, which is similar to Bhairavi but almost extinct today. The prevalence of this phrase in Group 3 suggests that the phrase inventory of Bhairavi may have evolved over time to absorb this phrase.

References

[1] Gopala Krishna Koduri, Sankalp Gulati, Preeti Rao, and Xavier Serra. Rāga recognition based on pitch distribution methods. Journal of New Music Research, 41(4):337–350, 2012.
[2] Kaustuv Kanti Ganguli and Preeti Rao. On the distributional representation of ragas: Experiments with allied raga pairs. Transactions of the International Society for Music Information Retrieval, 1(1), 2018.
[3] Preeti Rao, Joe Cheri Ross, Kaustuv Kanti Ganguli, Vedhas Pandit, Vignesh Ishwar, Ashwin Bellur, and Hema A. Murthy. Classification of melodic motifs in raga music with time-series matching. Journal of New Music Research, 43(1):115–131, 2014.
[4] V. Gowri Rammohan. An Insight Into the Lives and Works of the Musical Trinity: Tyagaraja, Muthuswamy Dikshitar, & Syama Sastry. Andhra University, 2006.
[5] T. V. Subba Rao. Ragas of the Sangita Saramrta by King Tulaja of Tanjore. Music

Academy, 1993.

A Wrong Chord or a Cool Chord? Information Theory May Predict Musical Surprise but Not Necessarily Pleasantness

Ling Qi (Georgia Institute of Technology); Claire Arthur (Georgia Institute of Technology)

Understanding why some musical surprises feel pleasing while others don't is central to theories of hedonic reward. Berlyne (1974) proposed that pleasantness arises from a balance between predictability and surprise. Information-theoretic (IT) models like IDyOM and PPM operationalize this idea by quantifying surprisal (IC) and uncertainty (Entropy) from learned probability distributions of musical events. Recent studies, including Cheung et al. (2019), report that IC-entropy interactions can predict perceived pleasantness, implying that musical events with similar IC and entropy should elicit comparable pleasure. To formally test this implication, we introduce a novel stimulus-generation approach. For each chord in a sequence, we swap it with every other chord in the vocabulary and recalculate IT values. We identify swappable pairs of sequences that differ by a single chord with IC and Entropy held constant at that position. Participants listen to the original and swapped versions and provide continuous pleasantness ratings for each chord. In Study 1, we trained a PPM model on the CoCoPops-Billboard corpus and selected 16 swappable pairs spanning the IC–Entropy space. Data from 124 participants showed that pleasantness ratings diverged significantly more at the swapped chord position than elsewhere, despite identical IT values. Neither harmonic function nor sensory dissonance predicted the degree of divergence. Overall, pleasantness ratings were negatively correlated with IC and positively correlated with entropy. Unlike Cheung et al. (2019), the IC-Entropy interaction did not reach significance, possibly due to differences in chord vocabulary and stimuli selection. Study 2 addresses these discrepancies by training an IDyOM model with the same corpus (Mcgill-Billboard), parameters, and filtering logic as Cheung et al. (2019), prior to performing chord swaps. In study 2, participants wore pupillometry glasses during the listening task, providing an implicit measure of perceived surprise alongside continuous pleasantness ratings. Data collection is ongoing. Taken together, IC and entropy are useful but incomplete predictors of musical pleasantness, sensitive to factors outside the information-theoretic framework. This work motivates the search for other perceptually-relevant covariates, such as tension and resolution strength, that can refine computational models of musical expectation.

Supporters

This conference is a part of the Digital Musicology Study Group of the International Musicological Society (IMS).

Supporters

We are very grateful for the generous support of the following partners:

- Universitätsbund Würzburg e.V.
- Vogel Stiftung Dr. Eckernkamp
- Institut für Musikforschung
- Zentrum für Philologie und Digitalität (ZPD)