

# Learning with Music Signals: Technology Meets Education

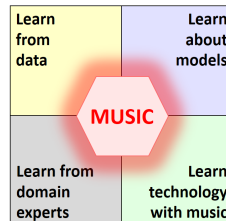
Meinard Müller<sup>\*1</sup>, Cynthia Liem<sup>\*2</sup>, Brian McFee<sup>\*3</sup>, and  
Simon Schwär<sup>†4</sup>

1 Universität Erlangen-Nürnberg, DE. [meinard.mueller@audiolabs-erlangen.de](mailto:meinard.mueller@audiolabs-erlangen.de)

2 TU Delft, NL. [c.c.s.liem@tudelft.nl](mailto:c.c.s.liem@tudelft.nl)

3 New York University, US. [brian.mcfee@nyu.edu](mailto:brian.mcfee@nyu.edu)

4 Universität Erlangen-Nürnberg, DE. [simon.schwaer@audiolabs-erlangen.de](mailto:simon.schwaer@audiolabs-erlangen.de)



---

## Abstract

Music information retrieval (MIR) is an exciting and challenging research area that aims to develop techniques and tools for organizing, analyzing, retrieving, and presenting music-related data. At the intersection of engineering, social sciences, and humanities, MIR relates to different research disciplines, including signal processing, machine learning, information retrieval, psychology, musicology, and the digital humanities. In Dagstuhl Seminar 24302, we explored advancing technology and education in these fields by examining learning from various angles, using music as a concrete application domain. Typically, learning in computer science brings to mind data-driven techniques like deep learning. While machine learning was crucial to the seminar, we aimed to go beyond a technical perspective, focusing on educational and pedagogical aspects. Specifically, we investigated how music can serve as a vehicle to make learning in signal processing and machine learning interactive and effectively communicated in interdisciplinary research and educational settings. In this report, we give an overview of the various contributions and results of the seminar. We start with an executive summary describing the main topics, goals, and group activities. Then, we give an overview of the participants' stimulus talks and subsequent discussions (listed alphabetically by the main contributor's last name) and summarize further activities, including group discussions and music sessions.

**Seminar** July 21–26, 2024 – <https://www.dagstuhl.de/24302>

**2012 ACM Subject Classification** Information systems → Music retrieval; Applied computing → Sound and music computing

**Keywords and phrases** Music Information Retrieval, Education, Signal Processing, User Interaction, Deep Learning

**Digital Object Identifier** 10.4230/DagRep.14.7.115

---

\* Editor / Organizer

† Editorial Assistant / Collector



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Learning with Music Signals: Technology Meets Education, *Dagstuhl Reports*, Vol. 14, Issue 7, pp. 115–152

Editors: Meinard Müller, Cynthia Liem, Brian McFee, and Simon Schwär



Dagstuhl Reports


Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

## 1 Executive Summary

*Meinard Müller (Universität Erlangen-Nürnberg, DE)*

*Cynthia Liem (TU Delft, NL)*

*Brian McFee (New York University, US)*

License  Creative Commons BY 4.0 International license  
© Meinard Müller, Cynthia Liem, and Brian McFee

This executive summary provides an overview of our discussions on advancing technology and education in music information retrieval (MIR) and related fields, summarizing the main topics covered in the seminar. We also describe the seminar’s group composition, overall organization, and activities. Finally, we reflect on the most important aspects of the seminar and conclude with future implications and acknowledgments.

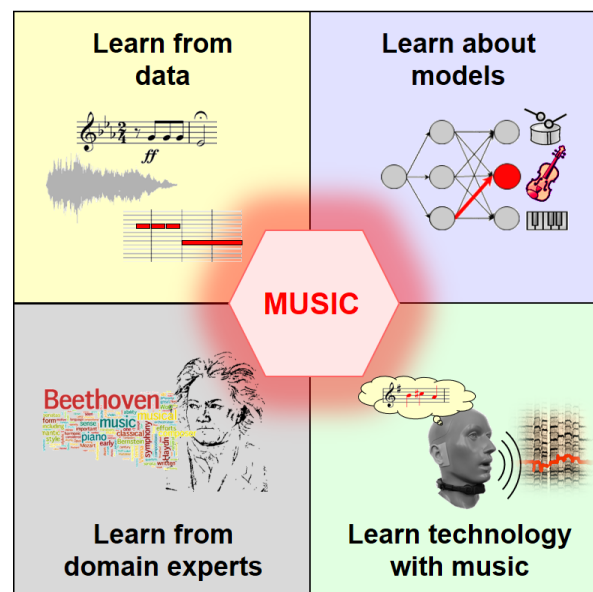
### Overview

In the last twenty years, the field of music information retrieval (MIR) has undergone rapid developments in terms of the problems considered, the methodology, and its applications. Using conceptually simple tasks and methods evaluated on small and idealized datasets in its beginnings, MIR now contributes to a wide range of concepts, models, and algorithms that extend our capabilities of accessing, analyzing, understanding, and creating music. Given the complexity and diversity of music, MIR research considers various aspects such as genre, instrumentation, musical form, melodic and harmonic properties, dynamics, tempo, rhythm, and timbre, to name a few. Furthermore, music is inherently multimodal, incorporating speech-like signals (e.g., singing), videos (e.g., of live performances), static images (e.g., scanned music scores), and text (e.g., lyrics and reviews). This wealth of data makes MIR an interdisciplinary and challenging field of research, which closely connects to technical disciplines such as signal processing, machine learning, and information retrieval, as well as mathematics, musicology, psychology, and the digital humanities.

Having the Dagstuhl Seminar 24302 titled “Learning with Music Signals: Technology Meets Education,” our objective was to advance technology and education in MIR and related disciplines using music as a challenging and instructive multimedia domain. Thinking of data-driven machine learning approaches, we discussed recent deep learning (DL) approaches and their ability to learn from training examples to make accurate predictions for previously unseen data. Furthermore, by learning from the experience of traditional engineering approaches, our aim was to better understand existing and build more interpretable DL-based systems (e.g., through integrating prior knowledge).

Beyond these technically oriented perspectives, an essential focus of our seminar was to approach the concept of learning from other angles, including pedagogical, educational, psychological, and user-centered viewpoints. We argue that music is an essential part of our lives that most people feel connected to. Therefore, music yields an intuitive entry point to support education in technical disciplines. In particular, we explored how music may serve as a vehicle to make learning and teaching signal processing and machine learning an interactive pursuit.

In all perspectives on learning, the question of reproducible research, including open access to data and software, is becoming increasingly important (so future insights can be built upon existing ones). As an overarching topic of our seminar, we discussed questions about open science and good scientific practice. This is a key issue, especially in higher education, where the open exchange of best practices and teaching materials significantly impacts the interdisciplinary and transnational education of the next generation of researchers.



■ **Figure 1** Learning with music signals. The figure illustrates the various perspectives on learning: advancing deep learning techniques, integrating traditional engineering knowledge for interpretability, collaborating with domain experts to understand music corpora, and using music to enhance interactive learning in technical disciplines.

In summary, in our Dagstuhl Seminar we approached and explored the concept of learning from different angles, see also Figure 1. Besides considering technological developments, the seminar equally addressed aspects of data and model understanding, transdisciplinary methodology and applications, science communication, and education.

## Participants and Group Composition

In our seminar, we had 28 participants from various locations around the world, including North America (eight participants from the United States), Asia (four participants from China, India, Japan, and South Korea), and Europe (16 participants from France, Germany, the Netherlands, Spain, Sweden, and the United Kingdom). Beyond geographic diversity, many participants had cross-cultural backgrounds and experiences. As naturally happens in international research fields, part of this comes from participants' work experiences in other countries and cultures than those of their country of birth. At the same time, several participants were first-generation for being in research/higher education, had second-generation migration backgrounds (meaning their parents were born in different countries, and often cultures, than themselves) – or, because of the nature of their affiliation, brought extensive experience in teaching with students from such backgrounds.

The seminar was not only international, but also highly interdisciplinary. While most of the researchers specialized in music information retrieval with a technical focus on signal processing and machine learning, we also had participants with backgrounds in musicology, human-computer interaction, science education, mathematics, computer vision, and other fields. This diversity stimulated cross-disciplinary discussions, bringing together experts from both technical and non-technical disciplines and highlighting opportunities for new

collaborations. Many participants had strong musical backgrounds, with some even having dual careers in engineering and music, leading to numerous social activities, including playing music together. We also aimed to foster variety in terms of seniority levels, with four Ph.D. students and three postdoctoral participants, as well as gender diversity, with 10 out of 28 participants identifying as female. More than half of the participants (16 out of 28) were attending Dagstuhl for the first time and expressed enthusiasm about the open and retreat-like atmosphere. In conclusion, by bringing together internationally renowned scientists and promising early-career researchers from different fields, our seminar provided support and encouragement for emerging talents on their academic paths.

## Overall Organization and Schedule

Dagstuhl Seminars are known for their flexibility and interactivity, encouraging participants to discuss ideas and raise questions rather than merely presenting research results. In keeping with this tradition, we set the schedule during the seminar, inviting spontaneous contributions focused on future-oriented content. This approach helped us avoid a conference-like atmosphere, where the emphasis is often on past research achievements. Furthermore, instead of sitting in rows, we removed all tables and arranged the seating in a half-circle of chairs, significantly enhancing eye contact and interaction among all participants.

After the organizers provided an overview of the Dagstuhl concept, we began the first day with self-introductions, where each participant shared their background, expectations, and wishes for the seminar. We then proceeded with brief stimulus talks, lasting 15 to 20 minutes, in which selected participants addressed critical questions related to the seminar's overall theme in a non-technical manner. Each talk smoothly transitioned into an open discussion among all participants, with the presenter acting as the moderator. These discussions were well-received and often extended for more than half an hour. The first day concluded with a brainstorming session on central topics reflecting the participants' interests, helping to shape the schedule and format for the following day.

On the subsequent days, we continued with short stimulus talks followed by long and intensive discussion rounds. We also incorporated group discussions, splitting into smaller groups to delve into specific topics in greater depth. The results and conclusions of these parallel group sessions, which lasted between 60 to 90 minutes, were then presented and discussed with the entire group. Additionally, we included panel-like elements featuring moderators, panelists, interviews, surveys, and game-like group activities. On the last day, we concluded the seminar with a session we called "self-outroductions," where each participant presented their personal view on the seminar's results. In summary, thanks to excellent group dynamics and a fair distribution of speaking time, all participants had the opportunity to express their thoughts, effectively avoiding a monotonous conference-like presentation format.

In addition to scientific questions, our seminar also addressed the various challenges that younger colleagues typically face when establishing their research groups and academic curriculum at the beginning of their careers. As previously mentioned, many of our participants had cross-cultural backgrounds, either being born in Asian countries or as second-generation individuals raised in Western cultures. One of the highlights of our Dagstuhl Seminar was a panel discussion on the cross-cultural challenges in academia, especially for individuals with Asian roots living and working in Europe or the US. This deeply personal and enlightening event was facilitated by Dagstuhl's unique environment, which fosters trust and mutual understanding.

While working in technical engineering disciplines, most participants also had a strong background and interest in music. This versatility significantly enriched the seminar's atmosphere, fostering cross-disciplinary interactions and sparking thought-provoking discussions. It also led to intensive joint music-making during breaks and evenings. A particular highlight was the Thursday evening concert organized by Cynthia Liem and Christof Weiß, where various ensembles formed by participants performed a wide variety of music, including classical, Irish folk, and jazz.

## Conclusions and Acknowledgment

At the Dagstuhl Seminar 24302, we used music as a motivating and tangible domain to explore different perspectives on learning. These perspectives stimulated conceptual discussions, laying the groundwork for future projects and academic curricula. We focused on how to teach and pass on new technologies to students, using music as a challenging application domain. With experts in MIR, signal processing, machine learning, software development, science education, and music sciences, our interdisciplinary seminar generated vibrant discussions and highlighted opportunities for new collaborations. Immediate outcomes, such as plans to share research data and software, also emerged from the discussions. We aimed to expose attendees, especially early-career researchers, to new ideas for designing academic curricula in computer science and beyond. Specific areas and topics addressed in this seminar included:

- Contextualized education
- Inclusive education
- Educational software systems
- Interactive software frameworks
- Science communication
- Transdisciplinary methodology and collaborative research
- Computational musicology
- Human-in-the-loop systems for music processing
- Data-driven machine learning for MIR
- Explainable deep learning for MIR
- Integration of musical knowledge
- Hybrid models for MIR
- Differentiable models for MIR
- Data mining, acquisition, measurement, and annotation
- Data/annotation quality
- Data accessibility and copyright issues
- Open science
- Reproducible and sustainable research
- Academic integrity and good scientific practice

Besides the scientific aspect, the social aspect of our seminar was equally important. We hosted an interdisciplinary, international, and interactive group of researchers, consisting of current and future leaders in our field. Many participants were visiting Dagstuhl for the first time and praised the open and inspiring setting. The group dynamics were excellent, with many personal exchanges and shared activities. Some scientists expressed their appreciation for the opportunity to engage in prolonged discussions with researchers from neighboring fields, something often impossible at typical conferences. A standout feature of our seminar

was the interaction between younger researchers at the beginning of their academic careers and established researchers and educators. This facilitated a deeply enriching exchange between different generations, promoting mutual trust and understanding. The intensive dialogue between these groups was truly outstanding and highlighted the unique value of our seminar.

In conclusion, our expectations for the seminar were not only met but exceeded, particularly in terms of networking and community building. We want to express our gratitude to the Dagstuhl board for giving us the opportunity to organize this seminar, the Dagstuhl office for their exceptional support throughout the organization process, and the entire Dagstuhl staff for their excellent service during the seminar. In particular, we want to thank Heike Clemens, Andreas Dolzmann, Marsha Kleinbauer, Simone Schilke, and Christina Schwarz for their invaluable assistance in the preparation and organization of the seminar.

## 2 Table of Contents

### Executive Summary

*Meinard Müller, Cynthia Liem, and Brian McFee* . . . . . 116

### Stimulus Talks and Further Topics

Uncertainty Estimation for Music and Education  
*Vipul Arora* . . . . . 123

Dual Domain Beat Tracking  
*Ching-Yu Chiu, Lele Liu, Christof Weiß, and Meinard Müller* . . . . . 124

What is Music?  
*Roger B. Dannenberg* . . . . . 125

Using Analog Modular Synthesizers for Teaching Signal Processing and Machine Learning  
*Christian Dittmar* . . . . . 125

Some Thoughts on Music Education for Robots  
*Zhiyao Duan* . . . . . 126

SMART: A Symbolic Music Algorithm Resource Toolkit  
*Mark Gotham* . . . . . 126

Can Music Kickstart Programming Activities and Contribute to Programming Education?  
*Masataka Goto and Jun Kato* . . . . . 128

On the Reliability of Different Data Modalities for Expressive Performance Analysis  
*Patricia Hu* . . . . . 129

Operationalizing the Complexity of MIR-ML Models for Enhanced Understandability, Interpretability, and Learning  
*Jaehun Kim* . . . . . 130

Teaching Elsewhere: Lessons From Small Liberal Arts Colleges in the US  
*Katherine M. Kinnaird, Christopher J. Tralie, Timothy Tsai, and Jordan Wirfs-Brock* . . . . . 130

What Research Examples Should We Set to Future Generations in Times of Hype?  
*Cynthia Liem* . . . . . 131

Open Source, Foundations, and Learning  
*Brian McFee* . . . . . 131

An Interactive Music Game for Singing Education and Data Collection  
*Peter Meier, Simon Schwär, and Meinard Müller* . . . . . 132

Introducing the TISMIR Education Track: What, Why, How?  
*Meinard Müller, Simon Dixon, Mark Gotham, Preeti Rao, Bob L. T. Sturm, and Anja Volk* . . . . . 133

Human-AI Music Ensemble and its Application to Music Education  
*Juhan Nam* . . . . . 133

Designing for Creative Learning  
*Alex Ruthmann* . . . . . 134

Three Critical Perspectives on Differentiable Digital Signal Processing <i>Simon Schwär and Meinard Müller</i> . . . . .	135
Towards Hybrid Models for Music Generation <i>Sebastian Stober</i> . . . . .	138
MIR of the Future (will Mostly Involve AI-Generated Music) <i>Bob L. T. Sturm</i> . . . . .	138
Designing High-Impact Undergraduate Research Excursions in MIR <i>Timothy Tsai and Meinard Müller</i> . . . . .	140
Music-Lyrics Matching: A New Approach to Lyrics Creation <i>Changhong Wang</i> . . . . .	141
Learning from Multiple Versions of Music <i>Christof Weiß and Lele Liu</i> . . . . .	141
Hearing is Believing: Audio Previews for Publishing Scholarship about Sound <i>Jordan Wirfs-Brock</i> . . . . .	142
<b>Working Groups and Panels</b>	
Cultural Differences in Education Systems <i>Ching-Yu Chiu, Lele Liu, Alia Morsi Patricia Hu, Jaehun Kim, and Cynthia Liem</i>	143
Cultivating Expertise: Sustainable Approaches to Learning and Research in MIR <i>Alia Morsi, Lele Liu, Patricia Hu, Ching-Yu Chiu, and Cynthia Liem</i> . . . . .	144
Generally Applicable Challenges And Opportunities for Teaching Music Information Retrieval <i>Participants of Dagstuhl Seminar 24302</i> . . . . .	145
Symbolic Music Computing Tools <i>Participants of Dagstuhl Seminar 24302</i> . . . . .	146
Computational Music Understanding <i>Participants of Dagstuhl Seminar 24302</i> . . . . .	147
<b>An AI-Generated Theme Song for Dagstuhl Seminar 24302</b> . . . . .	148
<b>Participants</b> . . . . .	152



## 3 Stimulus Talks and Further Topics

### 3.1 Uncertainty Estimation for Music and Education

Vipul Arora (*Indian Institute of Technology Kanpur, IN*)

License © Creative Commons BY 4.0 International license  
© Vipul Arora

While the majority of research today focuses on developing highly accurate systems, uncertainty estimation takes a more modest approach by quantifying the inaccuracies of these systems. Uncertainty estimation means that, in addition to providing the regular output (classification, regression, etc.), the system also indicates its confidence in that output. This is beneficial for downstream decision-making tasks, particularly in critical applications such as healthcare, financial investments, and self-driving cars [1]. It is also useful for active learning to facilitate efficient data labeling [2, 3]. Furthermore, beyond decision-making, uncertainty estimation has been found to improve training by accounting for noise and outliers [4].

Music information retrieval involves many tasks that can benefit from uncertainty estimation. [3] utilizes uncertainty estimation to actively adapt melody estimation models, thereby enabling efficient editing of annotations. Audio samples or segments with high uncertainty in the estimated melody are selected for manual annotations, which are subsequently used to update the model through adaptation.

Uncertainty can arise from:

1. Data (*aleatoric uncertainty*), which may involve noise or ambiguities.
2. The model (*epistemic uncertainty*), which may result from model limitations or inaccuracies.

The combination of both is referred to as *predictive uncertainty*. Active learning can efficiently reduce epistemic uncertainty by providing additional training data for the model.

Another area in MIR that can benefit immensely from uncertainty estimation is music education. Uncertainty estimation is particularly useful for interactive applications [5], and music education involves teachers and students interacting with the system. We are working towards building a system to detect singing mistakes [6], where what is marked as a mistake depends on the teacher and the proficiency level of the student, which leads to uncertainty. One of our current projects focuses on detecting ornaments (*alankars*) in Indian Art Music. Here again, there is uncertainty regarding what gets marked as an ornament due to the inherent complexity of the task. The idea of quantifying uncertainty is useful in these situations.


The Dagstuhl Seminar allowed me to present the ideas mentioned above and provided an opportunity to interact with many scholars. Several scholars, including Meinard Müller and Hanna Lukashevich, showed keen interest in these ideas and recognized their potential. Brian McFee shared with me his work on uncertainty estimation for music. I had productive discussions with Roger Dannenberg about building interactive music applications for Indian art music. I also discussed music education and music therapy with Anja Volk. I invited Chris Tralie to give a talk on music similarity at my institute. Through my interactions with others, I learned about journals such as JOSS, TISMIR, and other venues where I can submit our work.

## References

- 1 Max-Heinrich Laves, Sontje Ihler, Tobias Ortmaier, and Lüder A. Kahrs. Quantifying the uncertainty of deep learning-based computer-aided diagnosis for patient safety. *Current Directions in Biomedical Engineering*, 5(1):223–226, 2019.
- 2 Nagarathna Ravi, Thishyan Raj, and Vipul Arora. TeLeS: Temporal lexeme similarity score to estimate confidence in end-to-end asr. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, tbd., 2024.
- 3 Kavya Ranjan Saxena and Vipul Arora. Interactive singing melody extraction based on active adaptation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32:2729–2738, 2024.
- 4 Maximilian Seitzer, Arash Tavakoli, Dimitrije Antic, and Georg Martius. On the pitfalls of heteroscedastic uncertainty estimation with probabilistic neural networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2022.
- 5 Ervine Zheng, Qi Yu, Rui Li, Pengcheng Shi, and Anne Haake. A continual learning framework for uncertainty-aware interactive image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(7):6030–6038, 2021.
- 6 Vipul Arora, Suraj Jaiswal, Akshay Raina, and Sumit Kumar. Automatic detection and analysis of singing mistakes for music pedagogy. *TechRxiv*, 2023.

## 3.2 Dual Domain Beat Tracking

*Ching-Yu Chiu (Universität Erlangen-Nürnberg, DE), Lele Liu (Universität Würzburg, DE), Christof Weiß (Universität Würzburg, DE), and Meinard Müller (Universität Erlangen-Nürnberg, DE)*

License  Creative Commons BY 4.0 International license  
 © Ching-Yu Chiu, Lele Liu, Christof Weiß, and Meinard Müller

Beat tracking is one of the most fundamental tasks in Music Information Retrieval, reflecting our ability to perceive, follow, and understand music. Existing studies have explored beat tracking in both the audio and symbolic domains. However, there is limited literature addressing the relationship or differences between how beat tracking systems perform in these two domains. We conducted initial experiments using a symbolic beat tracker employing MIDI files as input data, and an audio beat tracker with audio recordings as input data. From the preliminary experiments, we observed the potential benefits of fusing beat tracking in the two domains. However, several factors complicate the experiments, making it difficult to pinpoint where performance improvements originate. Specifically, our current experiments involve various factors, including different data representations (event-based representation vs. frame-based representation), different model architectures (convolutional recurrent neural network vs. bidirectional long-short term memory network), and potential transcription errors when obtaining MIDI files from audio recordings. Additionally, further analysis is needed to quantify the differences between beat trackers in the two domains.

At the Dagstuhl Seminar, we presented our preliminary results, discussed our observations, and explored ways to refine the scope of the study. Participants offered suggestions on how to combine beat tracking in the two domains and recommended new directions for further investigation.

### 3.3 What is Music?

*Roger B. Dannenberg (Carnegie Mellon University – Pittsburgh, US)*

License © Creative Commons BY 4.0 International license  
© Roger B. Dannenberg

We have seen amazing new techniques for music composition and synthesis, but there is strong evidence that important music concepts and abstractions are not being learned by today’s systems. Simply put, machines are learning to imitate style without a deep understanding of what lies beneath the style and what could be altered or applied in other contexts. In contrast, humans have radically transformed music multiple times throughout history and across cultures. Even if we set aside the avant-garde and consider only music that is almost universally admired, composers consistently invent new kinds of music. This is one of our greatest challenges: to discover the principles that “define” music. I believe these principles are not just a set of commonalities such as “music has pitch and rhythm,” but are much deeper, having more to do with temporal and tonal structures interacting with human perception and cognition.

### 3.4 Using Analog Modular Synthesizers for Teaching Signal Processing and Machine Learning

*Christian Dittmar (Fraunhofer IIS – Erlangen, DE)*

License © Creative Commons BY 4.0 International license  
© Christian Dittmar

Over the last two decades, the community of modular synthesizer enthusiasts has grown considerably worldwide. For me personally, the COVID-19 pandemic was the trigger to revisit this sub-culture and become fascinated with it. Dabbling with modular synths appeals to those with an interest in electronic music, audio signal processing, electrical engineering, industrial design, and elitist nerdism, spiced with a collector’s frenzy. In the demo session of this Dagstuhl Seminar, I set up three racks with modular synthesis modules to demonstrate to the participants that these devices can be useful tools for teaching undergrad students basic concepts in Signal Processing and Machine Learning. More specifically, I presented three live experiments that could be achieved with quick re-wiring of the systems:

- Auralization of the sampling theorem by running a sinusoidal oscillator through a sample-and-hold module.
- Manual rotation of a phase-shifted signal pair in the two-dimensional plane using a weight matrix module and an oscilloscope.
- Approximation of a cosine signal from a triangle signal using a simple neural network made up of weight matrix and hyperbolic tangent modules.

### 3.5 Some Thoughts on Music Education for Robots

*Zhiyao Duan (University of Rochester, US)*


License  Creative Commons BY 4.0 International license  
© Zhiyao Duan

From cyberspaces to physical spaces, artificial intelligence (AI) is transforming every aspect of human society. Interacting and collaborating with robots that possess human-level intelligence is no longer just a dream but something very likely to occur in the next decade. In addition to helping humans complete mundane or dangerous tasks such as cleaning and searching, becoming human companions is also highly desirable. This requires robots to understand human civilizations and social norms. Music is considered one of the highest forms of human ingenuity and plays an important role in society. Teaching robots to read, play, and compose music will not only advance robotics and AI research but also benefit human society as a whole.

In my stimulus talk, I shared some of my thoughts on why, what, and how to teach music skills to robots. While MIR has made significant progress in cyberspaces toward machine musicianship, much work remains to be done in physical spaces, particularly with instrument-playing robots. I reviewed some existing work on musician robots and their limitations, and then proposed several research directions.

### 3.6 SMART: A Symbolic Music Algorithm Resource Toolkit

*Mark Gotham (Durham University, GB)*

License  Creative Commons BY 4.0 International license  
© Mark Gotham

Dagstuhl Seminars provide a very welcome opportunity to consider work in progress, conversation-stimulating topics, and discipline-wide issues. They are particularly useful for topics that require broad “buy-in” from a range of practitioners in the field. As such, they present the perfect opportunity to advance a discussion about a coordination effort I have been informally engaging in with interested parties for a long time. Some high-level observations will help provide the context and motivation for this topic:

- Music computing is not a large field. There are wonderful, enthusiastic practitioners, but relatively few of us compared to more populous fields.
- Relatedly, music computing is somewhat disparate, with practitioners spread out across the world, and sub-fields like MIR, music theory, and music psychology having rather limited interaction, despite their closely shared goals, tasks, and data.
- Educational resources can serve not only as material for preparing specific classes but also as a vehicle for consolidating the field (e.g., the function of textbooks as teaching tools and reference books for experts).
- Code libraries can likewise serve as another gathering point, supporting both newcomers with “how-to” guides and consolidating the field for expert practitioners.

While audio (and signal processing) are relatively well served with code libraries (e.g., McFee et al.’s *librosa*<sup>1</sup>) and pedagogical materials (e.g., Müller’s *Fundamentals of Music Processing*<sup>2</sup>), relatively little exists by way of computational resources for other parts of music computing, including the so-called “symbolic” music.

There are several promising initiatives that bring together algorithms in specific areas of symbolic music computing. These include *OMNISIA* (for pattern finding in Java), *synpy* (for rhythmic syncopation in Python), and *MIDI Toolbox*<sup>3</sup> (for melodic contour and more in MATLAB). While these efforts are welcome, they all originate from different research groups, are implemented in different programming languages, and in some cases, are no longer maintained.

Then there are larger libraries that could potentially bring these efforts together. At one time, *humdrum*<sup>4</sup> was a central point of reference. It continues to be used and maintained – an impressive 40 years later! However, there are downsides; for example, the “language-neutral” setup is commendable but not very inviting for newcomers who have computational expectations based on the landscape of 2024. Later, *music21*<sup>5</sup> emerged. First published in 2010/11, it too continues to be maintained and used. That said, the creator/maintainer recently made the explicit decision<sup>6</sup> that it is *not/no longer* intended to provide the holistic directory function stated here. Instead, it specifically invites sub-projects to operate independently, with or without music21 as a dependency. *Partitura* is arguably one such project, though it explicitly positions itself in contrast to this approach, suggesting that users “working in computational musicology” should adopt music21 instead.

As a result, students and researchers wishing to “get into” a topic often have to do a lot of “spade-work” to compare any new algorithm with existing work or even to make use of those existing algorithms. In short, given this state of affairs, and following conversations with the maintainers of these code libraries, it is clear that we need a new coordination effort to bring together the work cited above at a higher level. To begin addressing this, we propose a new code library that primarily serves to unify these algorithms. This library should be as user-friendly as possible, serving as a welcoming introduction for newcomers to the field and a valuable reference for those already active. Moreover, it should serve to:

- carefully credit previous work,
- “fill in the gaps” with new implementations where none are readily available, and
- expand into uncharted territory alongside ongoing, separately published research.

Dagstuhl provided a valuable opportunity to advance these ideas and gauge colleagues’ opinions and priorities. We leave Dagstuhl with:

- a clear and shared vision,
- the motivation to populate and maintain this resource,
- a substantially expanded core development team.

Finally, this positive effort complements more *specifically and exclusively* pedagogical initiatives, notably including the TISMIR education track<sup>7</sup> that features prominently elsewhere in this seminar.

<sup>1</sup> <https://librosa.org/>

<sup>2</sup> <https://audiolabs-erlangen.de/FMP>

<sup>3</sup> <https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/miditoolbox>

<sup>4</sup> <https://www.humdrum.org/>

<sup>5</sup> <https://github.com/cuthbertLab/music21>

<sup>6</sup> <https://groups.google.com/g/music21list/c/HF3tgkMvNWI/m/7vaIHR88BAAJ>

<sup>7</sup> <https://transactions.ismir.net/articles/10.5334/tismir.199>

### 3.7 Can Music Kickstart Programming Activities and Contribute to Programming Education?

*Masataka Goto and Jun Kato (AIST – Ibaraki, JP)*

**License** © Creative Commons BY 4.0 International license  
© Masataka Goto and Jun Kato

**Main reference** Jun Kato, Masataka Goto: “Lyric App Framework: A Web-based Framework for Developing Interactive Lyric-driven Musical Applications”, in Proc. of the 2023 CHI Conference on Human Factors in Computing Systems, CHI 2023, Hamburg, Germany, April 23-28, 2023, pp. 124:1–124:18, ACM, 2023.

**URL** <https://doi.org/10.1145/3544548.3580931>

At the Dagstuhl Seminar, we shared our recent experiences with four editions of the annual programming contest for “Lyric Apps” since 2020. Lyric apps, as we call them, are a new form of lyric-driven visual art that can render different lyrical content based on user interaction.

We first developed the “Lyric App Framework,” a web-based platform for creating interactive graphical applications that play music and display synchronized lyrics. After releasing this framework as the “TextAlive App API”<sup>8</sup>, we began holding annual programming contests in 2020<sup>9</sup>. The contests were open to the public and held online, with the winners announced at the “Magical Mirai” exhibition featuring “Hatsune Miku.” Hatsune Miku is the most popular singing voice synthesis software/character, and we conducted the contests in collaboration with Crypton Future Media, Inc., the company behind Hatsune Miku. Thanks to Hatsune Miku’s popularity, our contests gained significant attention, receiving 159 lyric app submissions over four years. For each contest, we provided a dedicated set of musical pieces and lyrics with appropriate permissions for online streaming, allowing contestants to focus on development and add novel interactive capabilities to existing musical pieces.

We observed that our contests attracted a broader community than the typical creative coding community. One contestant mentioned that they usually did not (or could not) participate in such competitions, but their love for music motivated them to develop and submit a lyric app. While interest in creative coding culture is growing, not all programmers feel confident that they are “creative” enough to enjoy creative coding. Lyric apps allow these programmers to rely on existing musical pieces for the “creative” part, thus kickstarting their programming activities. Programming itself is a creative activity, and we were pleased to see that lyric apps provided opportunities for more people to unleash their creativity. Notably, several contestants reported in the post-contest questionnaire that they were indeed novices, in some cases without any prior experience in web application development. We believe that lyric app development could be a valuable component of computer science education curricula.


---

<sup>8</sup> <https://developer.textalive.jp>

<sup>9</sup> [https://magicalmirai.com/2023/procon/index\\_en.html](https://magicalmirai.com/2023/procon/index_en.html)

### 3.8 On the Reliability of Different Data Modalities for Expressive Performance Analysis

Patricia Hu (Johannes Kepler Universität Linz, AT)

License  Creative Commons BY 4.0 International license  
© Patricia Hu

Expressive performance is an integral part of music-making. Expert performers shape various parameters such as tempo, timing, dynamics, and articulation in ways that go far beyond the notated score, resulting in an expressive interpretation that emphasizes the dramatic, affective, and emotional qualities of the music, thereby engaging listeners.

Previous work has studied expressive performance parameters in both symbolic and audio modalities [1, 2]. While audio data captures the richness of timbre and dynamics, supporting a holistic understanding of performance, symbolic modalities offer a more precise, structured representation suitable for systematically studying nuanced performance parameters such as timing and articulation. Additionally, in bridging the two modalities, transcribed symbolic (MIDI) datasets have been released in recent years, among other uses, for performance analysis.

Given the availability of different data sources, questions and concerns regarding the variability and reliability of different modalities naturally arise. Specifically, it is important to identify which expressive dimensions and features can be reliably and consistently studied using which data modality. Understanding the advantages and limitations of each data source for different performance parameters will enable a more informed use of existing datasets, both in terms of their technical and musical potential application domains.

As a first concrete step, we conducted a preliminary study [3] to explore the musical quality of different transcribed piano performances. To this end, we proposed musically informed piano transcription evaluation metrics that measure performance aspects such as timing, articulation, harmony, and dynamics in a transcription.


**Acknowledgments.** The author acknowledges support by the European Research Council (ERC), under the European Union’s Horizon 2020 research and innovation programme, grant agreement No. 101019375 *Whither Music?*.

#### References

- 1 Maarten Grachten and Carlos Cancino-Chacón. Temporal Dependencies in the Expressive Timing of Classical Piano Performances. In *The Routledge Companion to Embodied Music Interaction*, pages 360–369, 2017.
- 2 Michel Bernays and Caroline Traube. Expressive Production of Piano Timbre: Touch and Playing Techniques for Timbre Control in Piano Performance. In *Proceedings of the Sound and Music Computing Conference (SMC)*, 2013.
- 3 Patricia Hu, Lukáš Samuel Marták, Carlos Cancino-Chacón, and Gerhard Widmer. Towards Musically Informed Evaluation of Piano Transcription Models. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2024.

### 3.9 Operationalizing the Complexity of MIR-ML Models for Enhanced Understandability, Interpretability, and Learning

*Jaehun Kim (SiriusXM/Pandora – Oakland, US)*

License  Creative Commons BY 4.0 International license  
© Jaehun Kim


After the successful reintroduction of deep neural network models, using larger models fueled by large-scale datasets has become the standard practice for achieving better results. Recently, very large foundational models pre-trained on massive datasets have gained popularity as building blocks for MIR applications, providing the power of big data and big models. However, this approach leads to immensely high model complexity, making it difficult for users to understand how the resulting model works. From an engineering perspective, the large black-box approach can be tempting because it reduces the effort of building such a big model from scratch, or because one might not have access to the necessary datasets in the first place. This high complexity of models results in fewer opportunities for practitioners to understand the system’s mechanisms, affecting the maintainability and reliability of larger systems that potentially host multiple such pre-trained models.

Alternatives for tackling this issue include model post-hoc explanation, de-biasing, distillation, and so on. Another alternative could be measuring the complexity of the model to understand how unreadable the system is and possibly provide an alternative perspective on assessing the “best” model by incentivizing interpretable models. However, measuring model complexity remains an open question that needs further study, both in the ML and MIR fields. There are several principled ways to measure complexity in simpler probabilistic models, such as logistic regression, but these methods can quickly become complicated as the model’s complexity increases.

During the Dagstuhl Seminar, we explored key works in this area and discussed how they could be applied to MIR-specific models, integrating diverse perspectives from musicology, machine learning, and music theory. Our ultimate aim was to identify classes of more interpretable yet expressive models, which could hopefully enable us to use statistical models as tools for learning from data.

### 3.10 Teaching Elsewhere: Lessons From Small Liberal Arts Colleges in the US

*Katherine M. Kinnaird (Smith College – Northampton, US), Christopher J. Tralie (Ursinus College – Collegeville, US), Timothy Tsai (Harvey Mudd College – Claremont, US), and Jordan Wirfs-Brock (Whitman College – Walla Walla, US)*

License  Creative Commons BY 4.0 International license  
© Katherine M. Kinnaird, Christopher J. Tralie, Timothy Tsai, and Jordan Wirfs-Brock

The model of the American “small liberal arts college” (SLAC) may be relatively unknown to those outside of the US, but this model of holistic education across many disciplines, with a low student-to-faculty ratio, offers a unique perspective on teaching and learning. In this talk, four professors from different SLACs around the US discussed the challenges and opportunities their roles provide and how they teach music information retrieval both inside and outside the classroom. The goal of this talk was to stimulate discussion and reflection on different models for equipping and training the next generation of researchers.



### 3.11 What Research Examples Should We Set to Future Generations in Times of Hype?

*Cynthia Liem (TU Delft, NL)*

**License** © Creative Commons BY 4.0 International license  
© Cynthia Liem

**Main reference** Patrick Altmeyer, Andrew M. Demetriou, Antony Bartlett, Cynthia C. S. Liem: “Position: Stop Making Unscientific AGI Performance Claims”, in Proc. of the Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024, OpenReview.net, 2024.

**URL** <https://openreview.net/forum?id=AIXUuLCuMe>

Noticing the current AI and generative AI hypes, the surrounding incentives, and the types of papers that make it through the noise and get lauded for their contributions, I am worried. Technical engineering innovations and the (sometimes) marginal improvement of benchmarks are often praised, while the question of whether these innovations serve a true purpose seems to garner much less attention.

Over the past few months, I have served as an ISMIR meta-reviewer, a PhD and student supervisor, and a thesis reading committee member. Simultaneously, I have engaged extensively in public discussions on the societal impacts of AI applications. In my supervision, publication efforts, and public engagement, I have explicitly focused on asking fundamental questions: Is the problem well-articulated? What type of insight would best serve the articulated problem? Often, the answers to these questions have little to do with large-scale AI or generative AI innovations.

At the same time, in my roles as a meta-reviewer and reading committee member, I frequently experienced run-ins with peers when I questioned the extent to which a piece of work was sufficiently grounded in long-term related research, and whether the presented innovation and its evaluation results actually aligned with the underlying problem statement and provided deeper insights. Typically, in such situations, I have been told that the innovation is “cool,” the work resulted in publications, and that other successful published works share similar characteristics. Addressing the actual underlying questions, I was told, would be out of scope or too difficult. Indeed, my students and I struggle significantly more to get our work academically acknowledged, while those who align with the hype more easily gain visibility and funding.

Is this what research is supposed to be about? Am I, in encouraging my students to be critical, inadvertently harming their careers? Or is this precisely what academics, particularly those affiliated with public institutions, should advocate for more strongly?

### 3.12 Open Source, Foundations, and Learning

*Brian McFee (New York University, US)*

**License** © Creative Commons BY 4.0 International license  
© Brian McFee


Open source software provides a common foundation for both research and education. Instead of building tools entirely from scratch, practitioners can leverage existing implementations as building blocks to rapidly compose sophisticated systems and solve complex problems. At the same time, open source software can serve as an educational reference point that students can directly inspect and, with perhaps some effort, understand or modify.

In recent years, so-called “foundation models” have become a common and powerful approach to rapidly developing data-driven systems, especially in situations where task-specific training data may be relatively scarce – which is often the case in MIR. A common work-flow is to use a previously trained model as a “black-box” feature extractor, effectively replacing the role of hand-crafted audio features. While this approach is undoubtedly effective for system development and “pushing the needle,” its pedagogical value remains unclear.

In my Dagstuhl presentation, I highlighted the parallels and differences between traditional open source software and foundation models in the context of MIR. My goal was to stimulate discussion around our use of large pre-trained models, data dependencies, and potential research directions to improve our understanding of how to best use and teach these methods.

### 3.13 An Interactive Music Game for Singing Education and Data Collection

*Peter Meier, Simon Schwär, and Meinard Müller (Universität Erlangen-Nürnberg, DE)*

**License**  Creative Commons BY 4.0 International license  
© Peter Meier, Simon Schwär, and Meinard Müller

In previous work [1], we developed a prototype for an interactive music game called “Sing Your Way.” The prototype resembles a jump-and-run game and uses a gaming controller as the input device. Additionally, it incorporates the player’s singing voice to interact with the game world. For this, we estimate the pitch of a microphone signal in real time and use it to control parts of the game world. For example, by singing a note, a pitch line appears in the game that a player can jump on to add stair-like elements, allowing them to overcome obstacles and reach the end of a level. The game is deliberately kept simple, with a limited set of rules: players can move left, move right, jump, and sing notes, where pitch determines the vertical position in the game world.

Within this simple gaming environment, we consider two specific use cases: singing education and data collection for MIR research. First, the prototype serves as a tool for singing education, offering a fun and motivating environment for practicing singing techniques. Players are challenged with different gaming levels that include various voice-related tasks, such as singing the correct notes, intervals, chords, scales, or melodies. Combining gamification with educational objectives helps players train their ear and improve their vocal control in an engaging and playful manner. Second, the game can also serve as a valuable tool for collecting data on vocal performance and singing style. By capturing detailed information about players’ pitch accuracy, timing, and vocal range, educators and researchers can analyze trends and patterns in singing proficiency. This data can contribute to research in music education and assist in developing new, personalized methods for interactive vocal training and assessment.

At the Dagstuhl Seminar, we presented a demonstrator of our game and invited feedback and discussions on several topics: user experience, educational impact, technical improvements, potential collaborations, and data privacy. By actively playing our music game prototype, the Dagstuhl Seminar participants offered diverse perspectives and ideas on game level design for music education and MIR research.

#### References

- 1 Peter Meier, Simon Schwär, Gerhard Krump, and Meinard Müller. Evaluating Real-Time Pitch Estimation Algorithms for Creative Music Game Interaction. In *Proceedings of INFORMATIK 2023*, Gesellschaft für Informatik e.V., pages 873–882, Bonn, Germany, 2023.

### 3.14 Introducing the TISMIR Education Track: What, Why, How?

Meinard Müller (Universität Erlangen-Nürnberg, DE), Simon Dixon, Mark Gotham (Durham University, GB), Preeti Rao, Bob L. T. Sturm (KTH Royal Institute of Technology – Stockholm, SE), and Anja Volk (Utrecht University, NL)

**License** © Creative Commons BY 4.0 International license

© Meinard Müller, Simon Dixon, Mark Gotham, Preeti Rao, Bob L. T. Sturm, and Anja Volk

**Main reference** Meinard Müller, Simon Dixon, Anja Volk, Bob L. T. Sturm, Preeti Rao, Mark Gotham:

“Introducing the TISMIR Education Track: What, Why, How?”, *Trans. Int. Soc. Music. Inf. Retr.*, Vol. 7(1), pp. 85–98, 2024.

**URL** <https://doi.org/10.5334/TISMIR.199>

The Transactions of the International Society for Music Information Retrieval (TISMIR) offers a new education track for articles with a tutorial-style approach, focusing on MIR research and practical applications. This track covers a wide range of topics, including specific techniques, fundamental principles, and practical aspects, reflecting the diverse interests of the MIR community. We published an editorial to introduce this new track, highlighting the key characteristics of educational articles and exploring why the music domain may provide an intuitive and motivating setting for education across various levels and disciplines [1].

At the Dagstuhl Seminar, we discussed the participants’ views on education, their teaching experiences in MIR-related fields, and the needs of their students. Additionally, we reviewed their role in education and the recognition they receive for educational activities in their respective institutions. We believe that writing educational articles benefits both authors and readers by sharing expertise, enriching the knowledge base, and fostering innovation. It also provides researchers and PhD students with opportunities to refine resources, reflect on principles, and receive expert feedback. We discussed incentives for writing educational articles, as well as barriers that may prevent one from doing so, and explore the potential of educational articles in academia. Finally, we explored how to craft effective educational articles for the TISMIR educational track, laying the groundwork for a broader discourse on education within MIR and beyond.

#### References

- 1 Meinard Müller, Simon Dixon, Anja Volk, Bob L. T. Sturm, Preeti Rao, and Mark Gotham. Introducing the TISMIR education track: What, why, how? *Transaction of the International Society for Music Information Retrieval (TISMIR)*, 7(1):85–98, 2024.

### 3.15 Human-AI Music Ensemble and its Application to Music Education

Juhan Nam (KAIST – Daejeon, KR)

**License** © Creative Commons BY 4.0 International license

© Juhan Nam

The human-AI music ensemble is a fascinating topic that encompasses various MIR tasks such as automatic music transcription, score following, and automatic accompaniment. These MIR tasks pose significant technical challenges, including real-time processing, robustness to room acoustics and interfering sources, and on-the-fly adaptation to human performance. At the same time, the human-AI music ensemble must consider numerous human factors for musical interaction beyond the audio domain. For example, the AI performance

system should recognize musical cues from human gestures, handle human performance errors, and provide visual feedback and presence for communication.


In this talk, we described the technical components and challenges of the human-AI music ensemble and highlighted recent advances in the MIR community. We also presented cases of the human-AI music ensemble in real stage performances. Then, we shared ideas and discussed issues when applying these outcomes to music education settings in public schools. Finally, we envisioned future directions for multimodal integration with language models to achieve further improvements.

### References

- 1 Roger B. Dannenberg. An On-line Algorithm for Real-Time Accompaniment. In *Proceedings of the International Computer Music Conference (ICMC)*, 1984.
- 2 Lorin Grubb, Roger B. Dannenberg. A Stochastic Method of Tracking a Vocal Performer. In *Proceedings of the International Computer Music Conference (ICMC)*, 1997.
- 3 Akira Maezawa, Kazuhiko Yamamoto. MuEns: A Multimodal Human-Machine Music Ensemble for Live Concert Performance. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2017.
- 4 Yuen-Jen Lin, Hsuan-Kai Kao, Yih-Chih Tseng, Ming Tsai, Li Su. A Human-Computer Duet System for Music Performance. In *Proceedings of the ACM International Conference on Multimedia*, 2020.
- 5 Carlos Cancino-Chacon, Silvan Peter, Patricia Hu, Emmanouil Karystinaios, Florian Henkel, Francesco Foscarin, Nimrod Varga, and Gerhard Widmer. The ACCompanion: Combining Reactivity, Robustness, and Musical Expressivity in an Automatic Piano Accompanist. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-23)*, 2023.
- 6 Jiyun Park, Sangeon Yong, Taegyun Kwon, Juhan Nam. A Real-Time Lyrics Alignment System Using Chroma and Phonetic Features for Classical Vocal Performance. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024.
- 7 Taegyun Kwon, Dasaem Jeong, Juhan Nam. Towards Efficient and Real-Time Piano Transcription Using Neural Autoregressive Models, *Computing Research Repository (CoRR)*, abs/2404.06818, 2024.

## 3.16 Designing for Creative Learning

Alex Ruthmann (New York University, US)

License  Creative Commons BY 4.0 International license  
 © Alex Ruthmann  
 URL <https://musedlab.org/>

In my Dagstuhl Seminar talk, I explored the possibilities of creative learning design through the work of the NYU Music Experience Design Lab (MusEDLab). The lab’s research focuses on developing collaborative web applications and interactive systems that lower barriers to creative expression and sustain musical engagement by fostering curiosity and inquiry. The showcased applications embody these principles and demonstrate the lab’s innovative approaches.

Participants were invited to consider how Music Information Retrieval (MIR) can enhance these methodologies and support creative learning across diverse cultures and communities globally. The philosophical foundations discussed included the balance between design processes of iteration and ideation, the interplay between freedom and friction, and

the contrasts between Eastern and Western philosophies. Through these discussions, participants were encouraged to explore how MIR and interactive interface design can create open, creative spaces for music education, promoting sustained engagement and learning.

In connection with this presentation, I want to further summarize some related work. The book [1] outlines a musical journey through Scratch, an accessible programming environment with rich media features, including music and sound. It presents a series of independent musical projects, guiding readers through the process of creating each one. Along the way, readers explore coding techniques and algorithmic music processes. The interactive projects encourage music-making through play and composition. Designed for beginner coders and musicians, the book introduces programming and musical concepts as needed, with projects ranging from basic to advanced. Each project is self-contained, allowing readers to complete them in any order.

While much research has explored Scratch in the context of children creating games and interactive environments, little has focused on its music-specific functionality. Music and sound are central to children’s lives, yet creating music in coding environments often requires deep knowledge of music theory and computing. In [2], we examine the affordances and constraints of Scratch 2.0 for music-making. We analyze the limitations of its music and sound blocks, discuss bottom-up music programming, and break down the creation of a simple drum loop. We argue that current block designs may hinder meaningful music-making for children, Scratch’s core users. Additionally, we touch on the history of educational music coding languages, reference existing Scratch projects, compare Scratch with other music programming tools, and propose new block designs to foster more accessible music creation.

Finally, in [3], we discuss the development of “Sound Thinking,” an interdisciplinary general education course offered by the Departments of Computer Science and Music. The paper focuses on the student outcomes we aim to achieve and the projects designed to help students reach those outcomes. It explains our transition from a web-based environment using HTML and JavaScript to Scratch and discusses how Scratch’s “musical live coding” capability can more effectively reinforce these concepts.

## References

- 1 Andrew Brown and S. Alex Ruthmann. *Scratch Music Projects*. Oxford University Press, 2020.
- 2 William Payne and S. Alex Ruthmann. Music Making in Scratch: High Floors, Low Ceilings, and Narrow Walls? *Journal of Interactive Technology and Pedagogy*, 15, 2019.
- 3 S. Alex Ruthmann, Jesse M. Heines, Gena R. Greher, Paul Laidler, and Charles Saulters II. Teaching computational thinking through musical live coding in scratch. *Proceedings of ACM Technical Symposium on Computer Science Education (SIGCSE)*, pages 351-355, 2010.

## 3.17 Three Critical Perspectives on Differentiable Digital Signal Processing

*Simon Schwär and Meinard Müller (Universität Erlangen-Nürnberg, DE)*

License © Creative Commons BY 4.0 International license  
© Simon Schwär and Meinard Müller

Differentiable Digital Signal Processing (DDSP) is an umbrella term for the concept of including fixed signal processing components in deep learning models, requiring all building

blocks to be differentiable with respect to their input. This way, gradients can be back-propagated through the entire system, for example to train a neural network that controls the parameters of a fixed synthesizer. From the more general perspective of *hybrid* or *model-based* deep learning [1], this paradigm has two main advantages. First, the parameters of DDSP components are inherently interpretable. As opposed to *end-to-end* approaches, which tend to learn opaque representations of the training data, these parameters can be directly analyzed and modified with a lower susceptibility for confounding factors and overfitting. Second, incorporating domain knowledge (for example, about the physical properties of the sound generation mechanism) directly into the system architecture can be used as an inductive bias towards a computationally efficient model. Conversely, end-to-end approaches are often intentionally over-parametrized, resulting in unnecessary computational complexity.

Intuitively, it would be desirable to train a DDSP model with an unsupervised *analysis-by-synthesis* approach, i.e., by learning to estimate parameters for generating an output signal that is as close as possible to a given target signal. However, recent approaches using DDSP for various audio processing tasks [2, 3, 4, 5, 6, 7, 8, 9] typically require additional input information about the target or specifically designed loss terms to converge to meaningful solutions. While various work-arounds for training specific DDSP models have been proposed, we conjecture that fundamental issues may prevent fully unsupervised learning to be successful in estimating meaningful control parameters. At the Dagstuhl Seminar, we discussed the strengths, weaknesses and possible use cases of DDSP and hybrid or model-based deep learning in general, based on three different perspectives on this problem:

**1. DDSP as part of the loss function.** In many scenarios, a neural network is trained to output the parameters for a fixed DDSP model. For this, the gradient of a loss function between DDSP output and a target signal is first back-propagated through the fixed components, so that the loss for the neural network can equivalently be expressed as the composition of loss function and DDSP model. This view has recently led to new insights into training DDSP-based systems, and it has been shown that the “loss landscape” for individual parameters is often highly irregular, especially for parameters that control the frequency of synthesis components [10, 11, 12]. Further research is required into which combinations of loss functions and DDSP components result in favorable or problematic loss landscapes.

**2. DDSP as part of the neural network.** Some of the success of deep neural networks for complex tasks has been attributed to the high dimensionality of the loss landscape [13, 14]. While it is often inadequate to translate geometric intuitions to high-dimensional spaces, restricting individual parameters inside the network to have a specific physical meaning may undermine these benefits, since for DDSP components only very specific parameter combinations lead to an appropriate output. On the other hand, some restricting neural network components, like convolutional layers (inducing a bias towards local relationships in the data), are highly successful in a vast amount of application scenarios. This raises the question if there is a natural trade-off between the degree of restriction and learning success and which kinds of model biases are good choices for music and audio data.

**3. DDSP as an inverse problem.** An inverse problem is the task of determining from a set of *observations* their *cause* [15]. In technical terms, the goal is to estimate model parameters from measurements of the model’s output, which is closely related to the structure of many DDSP systems. It is well established that so-called *ill-posed* inverse problems are generally not easily solvable and require knowledge about the structure of each individual problem to obtain meaningful solutions in the parameter space [15]. While some

previous work has drawn initial parallels between DDSP and inverse problems [16, 17], it remains unclear to which degree various DDSP problems are ill-posed and whether methods from the field of inverse problems can be applied to improve training and results. In any case, this perspective warrants further research into specialized solutions for individual DDSP systems.

## References

- 1 Gaël Richard, Vincent Lostanlen, Yi-Hsuan Yang, and Meinard Müller. Model-based deep learning for music information research. *Computing Research Repository (CoRR)*, abs/2406.11540, 2024.
- 2 Jesse Engel, Lamtharn Hantrakul, Chenjie Gu, and Adam Roberts. DDSP: Differentiable digital signal processing. In *Proceedings of the International Conference on Learning Representations (ICLR)*, Virtual, 2020.
- 3 Jesse Engel, Rigel Swavely, Lamtharn Hanoi Hantrakul, Adam Roberts, and Curtis Hawthorne. Self-supervised pitch detection by inverse audio synthesis. In *International Conference on Machine Learning (ICML), Workshop on Self-Supervision in Audio and Speech*, Vienna, Austria, 2020.
- 4 Franco Caspe, Andrew McPherson, and Mark Sandler. DDX7: Differentiable FM Synthesis of Musical Instrument Sounds. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 608–616, Bengaluru, India, 2022.
- 5 Yusong Wu, Ethan Manilow, Yi Deng, Rigel Swavely, Kyle Kastner, Tim Cooijmans, Aaron Courville, Cheng-Zhi Anna Huang, and Jesse Engel. MIDI-DDSP: Detailed control of musical performance via hierarchical modeling. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2022.
- 6 Chin-Yun Yu and György Fazekas. Singing voice synthesis using differentiable lpc and glottal-flow-inspired wavetables. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 667–675, Milano, Italy, 2023.
- 7 Naotake Masuda and Daisuke Saito. Improving semi-supervised differentiable synthesizer sound matching for practical applications. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31:863–875, 2023.
- 8 Kilian Schulze-Forster, Clement S. J. Doire, Gaël Richard, and Roland Badeau. Unsupervised audio source separation using differentiable parametric source models. *Computing Research Repository (CoRR)*, abs/2201.09592, 2022.
- 9 Sungho Lee, Hyeong-Seok Choi, and Kyogu Lee. Differentiable artificial reverberation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30:2541–2556, 2022.
- 10 Joseph Turian and Max Henry. I’m sorry for your loss: Spectrally-based audio distances are bad at pitch. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), I Can’t Believe It’s Not Better Workshop*, Vancouver, Canada, 2020.
- 11 Ben Hayes, Charalampos Saitis, and György Fazekas. Sinusoidal frequency estimation by gradient descent. *Computing Research Repository (CoRR)*, abs/2210.14476, 2022.
- 12 Simon Schwär and Meinard Müller. Multi-scale spectral loss revisited. *IEEE Signal Processing Letters*, 30:1712–1716, 2023.
- 13 Yann N. Dauphin, Razvan Pascanu, Çağlar Gülçehre, Kyunghyun Cho, Surya Ganguli, and Yoshua Bengio. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *CoRR*, abs/1406.2572, 2014.
- 14 Yaim Cooper. The loss landscape of overparameterized neural networks. *CoRR*, abs/1804.10200, 2018.
- 15 Andreas Rieder. *Keine Probleme mit Inversen Problemen: eine Einführung in ihre stabile Lösung*. Vieweg Verlag, 2003.

- 16 Han Han, Vincent Lostanlen, and Mathieu Lagrange. Learning to solve inverse problems for perceptual sound matching. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32:2605–2615, 2024.
- 17 Ben Hayes, Charalampos Saitis, and György Fazekas. The responsibility problem in neural networks with unordered targets. In *The First Tiny Papers Track at ICLR 2023, Tiny Papers @ ICLR 2023, Kigali, Rwanda, May 5, 2023*, 2023.

### 3.18 Towards Hybrid Models for Music Generation

*Sebastian Stober (Otto-von-Guericke-Universität Magdeburg, DE)*

License  Creative Commons BY 4.0 International license  
© Sebastian Stober


Joint work of Jens Johannsmeier, Sebastian Stober

Generative modeling of music is a challenging task due to factors such as its hierarchical structure and long-term dependencies. Many existing methods, particularly powerful deep learning models, operate exclusively on one of two levels: the symbolic level, e.g., notes, or the waveform level, i.e., outputting raw audio without reference to any musical symbols. We argue that both approaches have fundamental issues that limit their potential for applications in computational creativity, particularly in open-ended creative processes. Specifically, symbolic models lack grounding in the reality of sound, while waveform models lack the level of abstraction necessary to make music creation more accessible.

At the Dagstuhl Seminar, we proposed a hybrid approach that integrates end-to-end modeling at both levels, combining their strengths while overcoming their respective limitations. While similar models already exist, they typically involve separate components that are only combined after training is complete. In contrast, we advocate for fully integrating both levels from the outset. We discussed the advantages and opportunities of this approach, as well as the challenges it presents and potential solutions. Our goal was to inspire other researchers to see the value in fully hybrid modeling of musical data going forward.

### 3.19 MIR of the Future (will Mostly Involve AI-Generated Music)

*Bob L. T. Sturm (KTH Royal Institute of Technology – Stockholm, SE)*

License  Creative Commons BY 4.0 International license  
© Bob L. T. Sturm

MIR of the future will look different than it does today, thanks to the coming deluge of AI-generated music content. The query-document paradigm of MIR is shifting to a prompt-generation paradigm, where music information retrieval has limited applicability – and indeed efficacy – in corpora projected to grow by “billions of [AI-generated] songs per year” [2]. With the proliferation of online commercial services providing music generation via prompting (e.g., [suno.com](https://suno.com) and [udio.com](https://udio.com)), or even generation-to-distribution pipelines (e.g., [boomy.com](https://boomy.com)), the sum total of human-crafted musical work is threatened to be swamped by AI-generated music – a “musicpocalypse,” echoing the “textpocalypse” predicted by Matthew Kirschbaum [1] in his *Atlantic* article about the impact of large language models (LLMs) like ChatGPT. As he writes, “Our relationship to the written word is fundamentally changing ... [These LLMs create] synthetic text devoid of human agency or intent ...



[and will lead to] a textpocalypse, where machine-written language becomes the norm and human-written prose the exception.”

In this stimulus talk, I presented two papers that deal with the “musicpocalypse” and its implications for MIR (and beyond). The first paper (rejected from ISMIR 2023) involved a collaboration among twelve authors from a variety of disciplines: engineering, musicology and ethnomusicology, sound studies, computational creativity, computer music, and composition. The authors include Bob L. T. Sturm, Ken Déguernel, Rujing S. Huang, André Holzapfel, Ollie Bown, Nick Collins, Jonathan Sterne, Laura Cros Vila, Luca Casini, David Alberto Cabrera Dalmazzo, Eric A. Drott, and Oded Ben-Tal. This paper playfully argues that a new kind of music studies is needed to address the impending flood of AI-generated music. It outlines six aspects for critical analysis and discusses each with reference to the contemporary AI music service *Boomy.com*: the service or company, the founders and employees, the use of the service, the users, the algorithms, and finally, the resulting music. After rejection, we posted this paper to SocArXiv<sup>10</sup>. An updated version has since been accepted to the 2024 AI Music Creativity conference<sup>11</sup>. We were also motivated to organize our own humanities-focused conference on the subject: “The First International Conference in AI Music Studies,” to be held Dec. 10-12, 2024, in Stockholm<sup>12</sup>.

While the paper above was criticized for being “more like the study of a business case,” having a weak technical analysis, and being irrelevant to ISMIR, we decided to write a paper clearly aligned with the technical focus of ISMIR. This paper (rejected from ISMIR 2024) is titled “AI Music: A New Frontier for Music Information Retrieval” and was written by Laura Cros Vila, Bob L. T. Sturm, Luca Casini, Anna-Kaisa Kaila, and David Dalmazzo. In this work (currently being revised for submission to TISMIR), we explore the implications of widespread AI-generated music for MIR. We present two concrete case studies centered on music collections generated by contemporary AI music services. The first case study (“music genre recognition”) analyzes a collection of music recordings generated by one AI music service to attempt to automatically uncover the music styles of the proprietary system. The second case study (“composer identification”) analyzes two collections of music recordings from different AI music services in relation to human-made music to try to identify the origin. Through these case studies, we identify many outstanding questions and promising avenues for future MIR research as it responds to the rising tide of AI-generated music.

The discussion of these two works during the week at Dagstuhl revealed that many participants are convinced of the novel and significant challenges posed to MIR and music in general by AI-generated music. The quality of the generated audio is so high that it is nearly competitive with what one hears on commercial radio. As an unplanned demonstration of that quality, we conducted an experiment when Christof Weiß asked me (on Wednesday afternoon) to generate musical audio from the prompt, “A medium-tempo funky song about music informatics researchers meeting in Germany with Irish-style melodic lines played by bass, drums, acoustic guitar, piano, vocoder, baritone sax, ...” Within 30 minutes, I had a finished 4m09s recording of a catchy piece of music – titled by the system, “Funk and Data in Deutschland” – which a group of us then transcribed and learned on Wednesday evening. We performed the song live at the Thursday evening concert to great applause. For more details on this experiment, see Section 5 of this Dagstuhl report.

---

<sup>10</sup> <https://osf.io/9pz4x>

<sup>11</sup> <https://aimc2024.pubpub.org>


<sup>12</sup> <https://boblsturm.github.io/aimusicstudies2024>

## References

- 1 Matthew Kirschenbaum. Prepare for the textpocalypse. <https://www.theatlantic.com/technology/archive/2023/03/ai-chatgpt-writing-language-models/673318/>, March 2023 (accessed September 4, 2024).
- 2 Valerio Valerdo. The sound of AI (conversations): Valerio Valerdo interviews Alex Mitchell. <https://youtu.be/iyTJF7b6BwE>, April 2022 (accessed September 4, 2024).

## 3.20 Designing High-Impact Undergraduate Research Excursions in MIR

*Timothy Tsai (Harvey Mudd College – Claremont, US) and Meinard Müller (Universität Erlangen-Nürnberg, DE)*

License  Creative Commons BY 4.0 International license  
© Timothy Tsai and Meinard Müller

In this talk, we described an experiment in designing high-impact research excursions for undergraduate students in Music Information Retrieval (MIR) as a culminating part of their undergraduate research projects. Education research has highlighted the importance of research experiences for undergraduates (REUs) in encouraging students, especially those from underrepresented minorities, to pursue graduate degrees. The Harvey Mudd College MIR Lab partnered with the Semantic Audio Processing Group at the International Audio Laboratories Erlangen (AudioLabs) to host joint workshops aimed at providing a compelling and impactful experience for undergraduates. At the end of their summer REU internship, the students travel to a host university to present their work, engage in technical discussions, learn about current research, and receive professional mentoring and feedback.

So far, we have conducted two joint workshops in 2023 and 2024 involving seven undergraduate students. These events were organized as two-day workshops with various components. First, students presented their work in roughly 15-minute presentations interleaved with and followed by 45-minute intensive discussions. These presentations stimulated deeper discussions not only on the research content but also on the conceptual elements of good scientific research practices and presentation styles. Second, the organizers gave scientific presentations related to the REU topics, bridging the gap to state-of-the-art techniques. Third, the undergraduate students had the opportunity to meet PhD students to discuss recent research directions and gain insights into their research careers and experiences as PhD students. Individual conversations with PhD students proved to be particularly impactful for the participants. Finally, joint meals, hiking, and other social interactions allowed students to meet informally with the organizers and other researchers. The workshop concluded with a day of sightseeing.

At Dagstuhl, we shared our observations and insights from these workshops and discussed with other participants the benefits to students, faculty, and the wider MIR community. In particular, brainstormed how the MIR community can educate, welcome, and encourage the next generation of researchers.

### 3.21 Music-Lyrics Matching: A New Approach to Lyrics Creation

Changhong Wang (Télécom Paris, FR)

License  Creative Commons BY 4.0 International license  
© Changhong Wang

Joint work of Changhong Wang, Gaël Richard

Alongside advancements in music generation, there is growing interest in lyrics or song generation within the music information retrieval community. Generating lyrics from scratch is non-trivial, as it requires consideration of not only text quality but also the match between music and lyrics. Prior efforts have sought to address both aspects simultaneously [1, 2], but have encountered difficulties primarily from a lexical standpoint, including grammatical correctness and semantic consistency. We propose *music-lyrics matching*, a new approach to creating lyrics by matching music with existing texts, leveraging the inherent relationships between music and lyrics.

#### References

- 1 Kento Watanabe, Yuichiroh Matsubayashi, Satoru Fukayama, Masataka Goto, Kentaro Inui, and Tomoyasu Nakano. A melody-conditioned lyrics language model. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, pp. 163–172, 2018.
- 2 Xichu Ma, Ye Wang, Min-Yen Kan, and Wee Sun Lee. AI-Lyricist: Generating music and vocabulary constrained lyrics. In *Proceedings of the ACM International Conference on Multimedia*, pp. 1002–1011, 2021.

### 3.22 Learning from Multiple Versions of Music

Christof Weiß (Universität Würzburg, DE) and Lele Liu (Universität Würzburg, DE)

License  Creative Commons BY 4.0 International license  
© Christof Weiß and Lele Liu

Main reference Lele Liu, Christof Weiss: “Utilizing Cross-Version Consistency for Domain Adaptation: A Case Study on Music Audio”, in Proc. of the The Second Tiny Papers Track at ICLR 2024, Tiny Papers © ICLR 2024, Vienna, Austria, May 11, 2024, OpenReview.net, 2024.

URL <https://openreview.net/forum?id=ZNg3YQQKWT>

The age of deep learning provides exciting possibilities and progress in MIR but also presents a number of challenges, including:

- The lack of robust, generalizable, and understandable models.
- The lack of large amounts of high-quality, annotated training data.

In several personal discussions at Dagstuhl, we addressed these problems and discussed some preliminary results [1] from a long-term research project at the University of Würzburg<sup>13</sup>, where we approach such challenges by exploiting cross-version datasets of classical music. These datasets contain, for the same musical work:

- Several modalities (score and audio).
- Several performances (interpretations, or arrangements).
- Several annotations (multiple experts).

In a broader context, these datasets also comprise:

<sup>13</sup><https://gepris.dfg.de/gepris/projekt/531250483>

- Several works by the same composer.
- Several composers from the same historical period.

We use these relationships to test generalization along different dimensions, such as generalizing to other versions of a work or to other works by a composer. Additionally, we can leverage cross-version relationships for learning musical concepts that remain constant or vary across versions in an unsupervised manner.


As a concrete example, we conducted experiments on multi-pitch estimation, addressing unsupervised domain transfer from one instrument (piano) to other, unseen instrumentations (singing/orchestra). We approached this using a teacher–student learning strategy, where we compare teacher annotations across versions (performances) in the target domain and use only consistent annotations as labels to train the student model. Enforcing such consistency across versions during student training helps to improve the challenging domain transfer. In the discussions at Dagstuhl, we received valuable feedback on these ideas and reflected on robustness across different music domains.

### References

- 1 Lele Liu and Christof Weiß, Utilizing Cross-Version Consistency for Domain Adaptation: A Case Study on Music Audio. *International Conference on Learning Representations (ICLR)*, Tiny Papers, Vienna, Austria, 2024

## 3.23 Hearing is Believing: Audio Previews for Publishing Scholarship about Sound

*Jordan Wirfs-Brock (Whitman College – Walla Walla, US)*

License  Creative Commons BY 4.0 International license  
© Jordan Wirfs-Brock

The dominant genre for disseminating academic scholarship, even for research that is about sound or is multimodal, is the text-based journal article. As a result, scholars who focus on sound and other sensory media are forced to translate their work into a thoroughly *non-interactive, non-sensory* medium. Yes, graphics and videos in supplementary materials can help, but the fact remains: In the journal article format, dynamic, lively, sensory, embodied, loud research – on sonification, on music information retrieval, on dance and movement, on how people converse and interact with the help of assistive technology – is stripped of its real-world meaning and flattened into black-and-white words on a page. The medium that researchers use to share their work is often an afterthought, an addendum to the “real” work of research, but it shouldn’t be: How knowledge is generated and how knowledge is communicated are intertwined. Sound, like each sensory modality, has its own unique epistemological possibilities: We can know things through sound (or touch, or taste) that we cannot know through sight.

To explore this tension, I presented a provocation: What would an audio-based journal for research about audio sound like? Building on Bernd Herzogenrath’s concept of sonic thinking, we proposed the idea of an audio preview, in the form of a one-minute audio clip, for communicating the sonic characteristics of research that can accompany traditional text-based journal articles. By encouraging researchers to communicate their work in an audio-only format, we hope to embrace, celebrate, and share the unique knowledge that can only be communicated through sound.

## 4 Working Groups and Panels

### 4.1 Cultural Differences in Education Systems

*Ching-Yu Chiu (Universität Erlangen-Nürnberg, DE), Lele Liu (Universität Würzburg, DE), Alia Morsi (UPF – Barcelona, ES), Patricia Hu (Johannes Kepler Universität Linz, AT), Jaehun Kim (SiriusXM/Pandora – Oakland, US), and Cynthia Liem (TU Delft, NL)*

**License**  Creative Commons BY 4.0 International license  
© Ching-Yu Chiu, Lele Liu, Alia Morsi Patricia Hu, Jaehun Kim, and Cynthia Liem

Thanks to the globalization of education, researchers, professors, and students now have numerous opportunities to work with international partners from different cultures, sparking diverse ideas, ways of thinking, and collaboration styles. However, many unexpected phenomena often go unrecognized or undiscussed. For instance, professors may find their Eastern students more shy, cautious, or passive compared to Western students during classes or communications. International students or researchers may notice that their supervisors have different expectations or emotional expressions compared to those in their home cultures. Since these phenomena are often not directly related to ongoing topics, tasks, or courses, they are rarely discussed or even acknowledged. Yet, they significantly influence individual thinking, perception, and behavior, impacting collaboration, communication, and educational outcomes. To address this, we held a panel session to openly discuss these observed phenomena. At the Dagstuhl Seminar, we explored the following aspects:

- Participants' experiences regarding expectations from their education systems, working environments, research communities, or family backgrounds.
- Participants' concerns when interacting with collaborators or students from different cultures.
- Potential scenarios that could facilitate mutual understanding and alleviate communication barriers.

During the discussion, we recognized the systematic differences between the education systems of various cultures, which are deeply embedded in their societies and not easily changed in the short term. The primary purpose of this panel discussion was to raise awareness of these issues and collaboratively build bridges to facilitate understanding and reduce communication barriers. To ensure that participants felt safe speaking openly, we agreed to keep the discussion contents anonymous and non-specific in this document. Ultimately, we gained a better understanding of cultural differences, recognized potential issues in our educational environments, and exchanged ideas to improve communication.

## 4.2 Cultivating Expertise: Sustainable Approaches to Learning and Research in MIR

*Alia Morsi (UPF – Barcelona, ES), Lele Liu (Universität Würzburg, DE), Patricia Hu (Johannes Kepler Universität Linz, AT), Ching-Yu Chiu (Universität Erlangen-Nürnberg, DE), Cynthia Liem (TU Delft, NL)*

License © Creative Commons BY 4.0 International license

© Alia Morsi, Lele Liu, Patricia Hu, Ching-Yu Chiu, and Cynthia Liem

Joint work of Participants of Dagstuhl Seminar 24302

Recent years have brought significant changes to many AI-related research areas, including music information research (MIR). One example is the increased demand for publications, which has affected the research climate by requiring researchers to respond quickly to keep pace with rapid developments. We have also observed growing popularity in certain research directions, along with shifts in research practices, including changes in the points of emphasis in paper writing. As early-career researchers, these changes have prompted us to reflect on how we can improve our personal career and skill development strategies, as well as consider how MIR might be affected overall. In this session, we held an open discussion and gathered responses to questions on the aforementioned themes. The remainder of this abstract summarizes the contributions from all attendees.

Regarding career development, we posed questions about the trajectory of gaining expertise, wondering whether and how our self-expectations should change during times of major shifts. We also sought opinions on what our development priorities should be during the PhD phase. Some participants shared anecdotes about their learning sources and strategies for navigating conferences to stay updated. One important strategy during a PhD is to develop the habit of discussing research with peers and colleagues. Another is to maintain a learning mindset, with the self-awareness and honesty to admit when something has not yet been fully understood. While each PhD journey is unique and requires developing in-depth knowledge in one's core topics, it is still useful to keep in mind the desired career path after the PhD to ensure that all necessary skills are developed to a sufficient level during the program.

The feeling of being an expert highly depends on personal self-evaluation, where it is common to fluctuate between moments of feeling like an expert and periods of the opposite. Our research endeavors naturally lead to this duality, as we become increasingly aware of how much we don't know. Once a researcher gains a grasp on a topic, they inevitably pose new questions and once again cross into areas where they feel like a beginner. This process is crucial for seeking new knowledge and fostering creativity. Moreover, we often become less impressed by the knowledge we have already gained, which can diminish our sense of accomplishment. An important intrinsic challenge, therefore, is maintaining motivation through these ups and downs. It is crucial to stay curious and connected with fellow researchers, including mentors and allies. Finally, understanding our natural predispositions for learning is key to setting realistic expectations and developing habits that align with our mental makeup.

A concerning side effect of the increasing quantity of publications is the difficulty of absorbing new works and insights in a timely manner, which, in the long run, could hinder the accumulation of knowledge within MIR. Some have expressed concern over the potential loss of nuance related to older techniques as earlier research is increasingly overlooked. This worry is exacerbated by a current trend in which industry research groups and industry-

funded academic collaborations dominate, leveraging their ability to develop larger models through somewhat brute-force approaches that, while yielding improved performance, shift focus away from less resource-intensive methods that may not perform as well.

The pursuit of improved metrics through computational force risks overshadowing the importance of building upon and refining established knowledge, potentially leading to the reinvention of existing solutions. Furthermore, it becomes harder to collectively advance the field when interesting insights regarding individual failure cases or system limitations are insufficiently highlighted (or even excluded) from papers. This can happen when global performance metrics are prioritized over providing musical context or understanding. Another side effect of data-hungry models is the surge in dataset development, often without careful consideration of the underlying data or task definitions, which can lead to problematic evaluations and poorly defined problems.

Finally, while there is growing awareness of the ethical implications of AI in the music sector, a significant gap remains in understanding how these changes will affect the broader human experience of music, including the pursuit of creativity and musicianship. We should aim to produce research that leaves a positive footprint and aligns with our values, while also remaining vigilant of the impact of changes beyond our control. A pressing example is AI music generation, given its current popularity. It is important for our research community to investigate the evolving landscape of music listening in light of AI-generated music entering streaming platforms. As the boundaries between human-composed and AI-composed music become increasingly blurred, there is a risk that AI-generated content may infiltrate future music datasets, compromising models' ability to learn the true diversity of real-world music. It is our duty to regularly reflect on the impact of our research.

### 4.3 Generally Applicable Challenges And Opportunities for Teaching Music Information Retrieval

*Participants of Dagstuhl Seminar 24302*

License  Creative Commons BY 4.0 International license  
© Participants of Dagstuhl Seminar 24302

Our aim for this breakout session was to discuss the unique challenges and opportunities of teaching music information retrieval in a college/university setting. The participants represented incredibly diverse teaching contexts, varying by size (from a dozen students up to 100), geographic location (US, Europe, Asia), the mix of undergraduates and graduates, and other factors. Despite these differences, we identified and worked through some common themes:

- **Interdisciplinary Nature:** We all face challenges with the interdisciplinary nature of the field, particularly when managing course prerequisites. Music information retrieval brings together signal processing, music theory, software engineering, and data science under one umbrella. A key learning goal is to aim for *literacy* and *precision of terminology* in every relevant discipline, regardless of each student's background.
- **Course Design Principles:** We identified a couple of important course design principles that are widely applicable. The first is “scaffolding,” which involves providing the necessary context for more complex tasks, breaking them down into smaller components, and offering sufficient training, starter code, or examples before allowing students to work independently. The second is “backwards design,” where you identify two to three key tasks, skills, and habits at the outset and structure the entire course around

them.

- **Teaching Neural Networks:** There was some discussion about how to teach students the important aspects of deep neural networks applied to our field, given the heavy computational requirements and the complexity of the software in a classroom setting. We identified differentiable digital signal processing (DDSP) as a good option, as certain versions only require 10 minutes of training data and networks with parameters “only” in the hundreds of thousands. However, there were other suggestions, such as overfitting to smaller examples or testing on smaller datasets, like subsets of the dataset “NSynth.” In this context, we also highlighted the usefulness of the resource on “Troubleshooting Deep Neural Networks.”<sup>14</sup>
- **Group Work:** Group work is often challenging to coordinate. To improve this, it helps to have multiple checkpoints for large projects. Additionally, each student in the group should take on the role of leader for at least one checkpoint, with no student serving as a leader more than twice.
- **Engagement:** Allowing students to choose their own music examples in MIR courses enhances engagement by making the content more relevant and enjoyable, encouraging deeper exploration of concepts through music they are passionate about. Therefore, we should allow students to choose the music examples whenever possible.

#### 4.4 Symbolic Music Computing Tools

*Participants of Dagstuhl Seminar 24302*

License © Creative Commons BY 4.0 International license  
© Participants of Dagstuhl Seminar 24302

The group session focused on developing a new software tool for symbolic music analysis, aiming to provide essential functionality and baselines. Key discussion points included the need for foundational algorithms for melodic contours and large-scale music analysis, such as extracting melody, structure, and chords from MIDI data. Participants emphasized the importance of structured, annotated data to improve AI, calling for the integration and crosslinking of multimodal metadata.

Debates on data formats highlighted the pros and cons of CSV, MusicXML, MIDI, and MEI, with consensus favoring the refinement of existing formats over creating new ones. Parsing and handling of input and output representations should ideally be managed by existing tools. The tool should accommodate diverse music data types and formats, aiming for minimal, task-specific representations that align with algorithm inputs.

Usability and adoption were key concerns, with user surveys suggested to identify different user groups and their needs. The discussion also addressed the interaction between audio and symbolic data, event-based vs. frame-based processing, and distinctions between performed and non-performed music. The importance of open-source software tools capable of handling multiple formats and cross-modal information was emphasized. Challenges such as transcription standards in non-Western music, minimal representations, and tasks like score alignment and lyric distribution analysis were also discussed. The ultimate goal is to create a versatile, user-friendly software tool that consolidates existing algorithms and formats, providing a solid foundation for symbolic music analysis.

---

<sup>14</sup><https://fullstackdeeplearning.com/spring2021/lecture-7/>



## 4.5 Computational Music Understanding

*Participants of Dagstuhl Seminar 24302*

License © Creative Commons BY 4.0 International license  
© Participants of Dagstuhl Seminar 24302

We began our discussion on music understanding with a working definition of music as “structured sound” but quickly realized this definition might be insufficient, as some music’s structure is only discernible through higher-order cognitive processing (e.g., John Cage’s 4’33”). Building on this idea, we differentiated between music that involuntarily evokes emotional or physiological responses (e.g., through regular beats) and music that requires active cognitive, social, or emotional processing to elicit similar responses. We introduced the terms “art” or “experimental” music for the latter and “visceral” or “non-experimental” music for the former, focusing our discussion on the latter as it seemed more promising for understanding how humans “understand” music.

Next, we distinguished between objective and subjective music understanding. Objective music understanding refers to measurable aspects of music, such as beats, frequencies, pitches, and structure, which are the primary focus of most Music Information Retrieval (MIR) algorithms and tasks. Subjective music understanding, as we defined it, goes beyond measurable aspects, raising the question of whether subjective understanding is merely the sum of all objective measurements – whether a listener’s response can be fully predicted by knowing all the objective features of a musical piece. We also considered factors like nostalgia, memory, and mood, which may significantly influence subjective music understanding.

Adopting a more technically oriented perspective, we continued our discussion on MIR tasks where both objective and subjective music understanding could be beneficial. We specifically focused on the evaluation of AI-generated music, noting that current loss functions are often inadequate because they fail to account for domain-specific features, such as musical structure. Enhancing both objective and subjective music understanding could lead to more targeted and effective loss functions and evaluation metrics. Overall, we agreed that most topics and tasks in MIR would benefit from further research in music understanding, as this would make metrics and evaluations more human-centric.

We concluded our discussion by exploring the frontiers in objective and subjective music understanding. Several key areas of interest were identified, including uncovering common semantic metaphors shared by humans when listening to music, estimating emotions from music descriptors and features, and building predictive models from first principles (e.g., using architectures like PredNet [1]). We observed that current perception research experiments might be overly simplistic in capturing the complexities of music, as they often rely on highly controlled stimuli. Finally, we emphasized that research in music understanding is inherently multidisciplinary and would benefit from increased collaboration between fields such as neuroscience, psychology, musicology, and computer science.

### References

- 1 William Lotter, Gabriel Kreiman, and David D. Cox. Deep predictive coding networks for video prediction and unsupervised learning. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*, Toulon, France, 2017.

## 5 An AI-Generated Theme Song for Dagstuhl Seminar 24302

*“A medium-tempo funky song about music informatics researchers meeting in Germany with Irish-style melodic lines played by bass, drums, acoustic guitar, piano, vocoder, baritone sax ...”*

This text prompt, originally scribbled on a whiteboard in the seminar room during a working session, set off an exciting musical journey that culminated in a spontaneous live performance at our Dagstuhl Seminar. With AI-generated music being a prominent topic in many of the stimulus talks and discussions, it was only natural to explore cutting-edge deep learning tools to create a theme song for the event. Bob Sturm, one of the Dagstuhl participants, took up the challenge, using Suno<sup>15</sup> and Udio<sup>16</sup>, two online platforms for personalized AI-generated music, to craft the song’s lyrics, musical themes, and final version, aptly named **“Funk and Data in Deutschland.”** These AI tools directly generated the music recordings for the song. The step-by-step process of crafting the song is outlined in more detail below. Inspired by the creation, we decided to bring the AI-generated music to the stage with a live performance. Christof Weiß, another Dagstuhl participant, transcribed the AI-generated track into a symbolic lead sheet, complete with lyrics and chords (Figure 2), and arranged parts for a horns section (Figure 3). After some short and spontaneous rehearsals, it all came together in a fun and lighthearted performance by 13 seminar participants during our traditional Dagstuhl Seminar concert on Thursday evening. The entire creative process, including intermediate stages, has been documented on a dedicated website<sup>17</sup>.

Asking Bob Sturm about the process of creating the song and using the AI tools alongside his manual intervention, he explained it as follows:

- Initially I prompted Suno with: *“A medium-tempo funky song about music informatics researchers meeting in Germany with Irish-style melodic lines played by bass, drums, acoustic guitar, piano, vocoder, baritone sax.”* It generated lyrics and two recordings, one titled “Funk and Data in Deutschland.”
- I was not too excited by either of these recordings, so instead tried prompting Udio with: *“A medium-tempo funky song about music information retrieval researchers meeting in Germany, featuring Irish-style melodic lines played by bass, drums, acoustic guitar, piano, vocoder, baritone sax, accordion, melodica, and other percussion.”* This generated two 32-second recordings, and a few verses of lyrics.
- Not yet fully satisfied, I tried a second time with the same prompt but specified the lyrics from the first verse of “Funk and Data in Deutschland” generated by Suno. This resulted in two additional 32-second recordings.
- I selected one of these recordings for Udio to extend, adding another verse from “Funk and Data in Deutschland.” I chose this output because it better suited the instrumentation and had a ska-like sound that fit our intended orchestration.
- I continued to expand the song in Udio, adding verses one at a time, until I had a full song of 3m49s duration.
- Afterward, I exported the recording into a digital audio workstation, where I trimmed part of the introduction and applied compression and equalization.

<sup>15</sup> <https://suno.com>

<sup>16</sup> <https://www.udio.com>

<sup>17</sup> <https://audiolabs-erlangen.de/resources/MIR/2024-DagstuhlThemeSong>

- Overall, the process from the initial Suno prompt to the final track took no more than 40 minutes.

The result was a generated audio file of an entertaining, light, and fairly simple standard pop song in D major, featuring a pop band setup (guitar, bass, drums, vocals), along with trumpets and saxophones, and additional instruments for effects such as accordion, melodica, and piano. From this audio file, Christof Weiß manually transcribed a lead sheet with lyrics and chords (Figure 2) as well as a horn section arrangement (Figure 3, typeset by Christian Dittmar). This material (audio and transcriptions) formed the basis for rehearsing the song in a roughly one-hour session, involving 13 seminar participants directed by Christof Weiß, with Peter Meier as the lead singer. The song's overall simplicity made it easy to rehearse; after clarifying the structure to the musicians, we quickly achieved a reasonable performance. This confirmed our assumption that music generation systems often produce straightforward, familiar musical structures that appeal to mainstream audiences. Finally, we performed the song at the evening concert, introducing the creation process and inviting the audience to follow along and even sing the lyrics.

## FUNK AND DATA IN DEUTSCHLAND

### [Intro A+B]

D – A – G – A (8x) ctd.

### [Verse]

We hittin' Deutschland  
Jammin' with the pros ... (oo-yeah!)  
Researchin' data  
Where the music flows  
Bass and drums gettin' funky  
Don't you know? ... (uh-huh)  
Acoustic guitar strummin'  
Let that rhythm go!

### [Intro A] (2x)

### [Verse 2]

Piano keys dancin' with the sax so sweet ... (oh-oh!)  
Vocoder singin'  
Movin' with the beat  
Research nerds groovin' in the German heat  
From cities clear to towns  
In the street

### [Chorus]

G  
Funk and data  
G D  
Feelin' so alive ... (so alive!)  
G  
Researchers gather  
G D G E D  
Watch 'em thrive  
G  
In the heart of Deutschland  
G D  
Where the rhythms drive  
G  
A melodic mission  
G D G E D  
We unite We jive! (oo-yeah!)

### [Intro A] (2x)

### [Verse 3]

Harmonic lines like an Irish dream  
(come on now!)  
Mix of cultures in a fluid stream  
MIR researchers form the team  
Got that funky fever  
Or so it seems!  
Got the Dublin soul with a Teutonic twist  
Information retrieval  
Can't resist

### [Bridge]

G  
Let the algorithms  
G D  
groove like this  
G  
In that German spirit  
G D G E D  
Find the bliss!

### 4 bars of ESKALATION over D7

### [Chorus]

G  
Funk and data  
G D  
Feelin' so alive ... (so alive!)  
G  
Researchers gather  
G D G E D  
Watch 'em thrive  
G  
In the heart of Deutschland  
G D  
Where the rhythms drive  
G  
A melodic mission  
G D G E D  
We unite We jive! (oo-yeah!)

### [Intro A] (4x) ... moving into church

■ Figure 2 Lyrics and Chord Sheet of “Funk and Data in Deutschland”.

## FUNK AND DATA IN DEUTSCHLAND

Composer: Suno.ai & Udio.ai Prompter: Bob Sturm / Arr: Christof Weiß  
Typist: Christian Dittmar

$\text{♩} = 150$     Intro A

Bb Trpts.    1.    2.

Eb Bari.

6    Intro B

10    Verse 1    Intro A  
16    4    16    4  
1x senza rep.

34    Verse 2    14    Chorus  
14

56    Intro A 2x  
(1x con rep.)    2

66    Verse 3    22    Bridge  
22

94    E7 Eskalation 8  
H7 Eskalation 8

Chorus

Intro A 4x (2x con rep.)

The image shows a musical score for a horn section, specifically for Bb Trumpets and Eb Baritone. The score is titled 'FUNK AND DATA IN DEUTSCHLAND' and is arranged by Christof Weiß. It features several sections: Intro A, Intro B, Verse 1, Verse 2, Verse 3, Chorus, Bridge, and Eskalation (E7 and H7). The tempo is marked as quarter note = 150. The key signature is three sharps (F#, C#, G#). The score includes various musical notations such as slurs, accents, and dynamic markings like 'ff'. There are also repeat signs and first/second endings. The score is divided into systems, with measures 6, 10, 34, 56, 66, and 94 marking the beginning of new sections.

■ Figure 3 Horn section arrangement for “Funk and Data in Deutschland”.

## Participants

- Vipul Arora  
Indian Institute of Technology  
Kanpur, IN
- Ching-Yu Chiu  
Universität Erlangen-Nürnberg,  
DE
- Roger B. Dannenberg  
Carnegie Mellon University –  
Pittsburgh, US
- Christian Dittmar  
Fraunhofer IIS – Erlangen, DE
- Zhiyao Duan  
University of Rochester, US
- Mark Gotham  
Durham University, GB
- Masataka Goto  
AIST – Ibaraki, JP
- Patricia Hu  
Johannes Kepler Universität  
Linz, AT
- Jaehun Kim  
SiriusXM/Pandora –  
Oakland, US
- Katherine M. Kinnaird  
Smith College –  
Northampton, US
- Cynthia Liem  
TU Delft, NL
- Lele Liu  
Universität Würzburg, DE
- Hanna Lukashevich  
Fraunhofer IDMT –  
Illmenau, DE
- Brian McFee  
New York University, US
- Peter Meier  
Universität Erlangen-Nürnberg,  
DE
- Alia Morsi  
UPF – Barcelona, ES
- Meinard Müller  
Universität Erlangen-Nürnberg,  
DE
- Juhan Nam  
KAIST – Daejeon, KR
- Alex Ruthmann  
New York University, US
- Simon Schwär  
Universität Erlangen-Nürnberg,  
DE
- Sebastian Stober  
Otto-von-Guericke-Universität  
Magdeburg, DE
- Bob Sturm  
KTH Royal Institute of  
Technology – Stockholm, SE
- Christopher J. Tralie  
Ursinus College – Collegeville,  
US
- Timothy Tsai  
Harvey Mudd College –  
Claremont, US
- Anja Volk  
Utrecht University, NL
- Changhong Wang  
Télécom Paris, FR
- Christof Weiß  
Universität Würzburg, DE
- Jordan Wirfs-Brock  
Whitman College –  
Walla Walla, US

