Agentic AI for Music Knowledge Representation: Towards the Automated Generation of IEEE 1599 Documents

Luca Andrea Ludovico
Laboratory of Music Informatics (LIM)
Department of Computer Science
Università degli Studi di Milano
Italy
e-mail: luca.ludovico@unimi.it

Davide Andrea Mauro
KreativInstitut.OWL
Universität Paderborn
Germany
e-mail: davide.mauro@uni-paderborn.de

Abstract*—IEEE 1599 is a multilayer format designed to represent and synchronize heterogeneous music information, including symbolic notation, audio recordings, metadata, and analytical descriptions. Despite its expressive power, the creation of IEEE 1599 documents remains a labor-intensive process that often requires the integration of data originating from multiple sources and processing pipelines. Recent advances in Artificial Intelligence (AI), particularly in the field of AI-based agents, offer new opportunities for automating and coordinating such tasks. This contribution discusses the potential role of specialized AI agents in the generation of IEEE 1599 documents. We identify three application scenarios that are particularly relevant to the automatic population of the *Logic layer* and the synchronization structures of the format. First, conversion agents may transform existing symbolic encodings, such as MusicXML and MEI (Music Encoding Initiative), into IEEE 1599 representations. Second, Optical Music Recognition (OMR) agents may extract symbolic information directly from digitized scores and generate structured music descriptions. Third, audio synchronization agents may automatically align symbolic and audio content, producing synchronization data required by the format. We argue that an agent-based architecture can orchestrate these heterogeneous processes through autonomous task execution, verification, and refinement, thus reducing human intervention while preserving data quality. The paper outlines a conceptual framework for such an architecture and discusses research challenges, including reliability, interoperability, provenance, and validation of generated content. We conclude by highlighting how agentic AI may contribute to the broader vision of intelligent music archives and digital cultural heritage systems.

Keywords- AI-based Agents; Multi-Agent Systems; IEEE 1599; Music Information Representation; Music Information Retrieval; Optical Music Recognition; Audio-Score Alignment; Digital Music Libraries.

I. Introduction

The representation of music information increasingly requires the integration of heterogeneous media, including symbolic notation, audio recordings, metadata, annotations, and analytical descriptions. As better detailed in the next section, the IEEE 1599 format [1] addresses this challenge through a multilayer representation model capable of establishing explicit relationships among different forms of music content.

Despite its flexibility and expressive power, the creation of IEEE 1599 documents remains a complex and largely manual process. Recent advances in Artificial Intelligence, and particularly in AI-based agents capable of autonomous task execution and coordination, suggest new opportunities for automating the generation of IEEE 1599 representations.

This paper explores how agentic AI may support the creation of IEEE 1599 documents by integrating information from heterogeneous music sources and processing pipelines. Although the discussion is grounded in the specific case of IEEE 1599, the proposed approach is intended to have broader applicability. The same principles can be extended to other domains where complex knowledge objects must be constructed by combining symbolic descriptions, multimedia resources, metadata, annotations, and synchronization information. In this sense, IEEE 1599 provides a particularly suitable test bed for investigating a more general problem: how agent-based architectures can orchestrate heterogeneous tools, validate intermediate results, and support the semi-automatic generation of structured, multilayer knowledge representations.

The remainder of the paper is organized as follows. Section II provides the necessary background on the IEEE 1599 standard and AI-based agents. Section III discusses the main challenges involved in the production of rich IEEE 1599 encodings, with particular attention to the effort required to create and synchronize heterogeneous materials. Section IV introduces an agent-based vision for IEEE 1599 document creation, focusing on symbolic format conversion, Optical Music Recognition (OMR), and automatic audio synchronization. Section V discusses open research challenges, including reliability, validation, interoperability, explainability, provenance, and ethical issues. Finally, Section VI concludes and indicates possible directions for future work.

II. Background

A. IEEE 1599

IEEE 1599 is an international standard designed to encode music-related information. It is an XML-based format designed to offer a comprehensive and multi-layered representation of music and music-related materials. It focuses on providing multiple representations of the same piece within a unique XML document.

IEEE 1599 can link multiple and heterogeneous representations of the same music piece through a structured approach that organizes information into six interconnected layers.

A key feature of IEEE 1599 is its synchronization mechanism, which allows different representations (such as many audio recordings, scores, and computer-driven performances) to be linked to the same logical structure. Thanks to the adoption of a data structure called the *spine*, heterogeneous information is not only available and organized within a unique XML document but also interconnected. Synchronization among different materials uses time-based references (e.g., for audio and video) or location-based references (e.g., for graphical content). This makes it possible to seamlessly switch between different formats or modalities and navigate interactively within a musical piece.

Providing an in-depth description of the IEEE 1599 format would go beyond the scope of this paper, especially since the standard is already well documented in the scientific literature; interested readers may refer to the seminal workby Baggi and Haus [2].

B. AI-Based Agents

In recent years, the rapid rise of AI (Artificial Intelligence) techniques has paved the way to generative models and agentic systems capable of autonomous reasoning, perception, and action. At the core of many recent agent architectures are Large Language Models (LLMs), which act as general-purpose reasoning engines able to interpret complex instructions, interact through natural language, maintain contextual information, and select actions in pursuit of a given goal. Unlike conventional prompt-response systems, AI-based agents are usually conceived as goal-oriented entities embedded in an execution loop: they observe the current state of the task or environment, reason about possible next steps, select or invoke suitable tools, evaluate intermediate results, and iteratively refine their behavior [3].

A defining characteristic of Agentic AI is autonomous planning, which allows an agent to decompose a high-level, often ambiguous, prompt into a structured sequence of executable steps [4]. This planning process may involve task decomposition, plan selection, external tool invocation, feedback integration, and re-planning when intermediate results are incomplete or inconsistent [5].

Another important component of agentic systems is memory. Beyond the short-term context provided by the prompt, agents may rely on episodic, semantic, or task-specific memory structures to store previous observations, intermediate results, user preferences, and reflections on past failures [6][7]. Memory mechanisms make it possible to preserve continuity across multi-step processes and to support more adaptive behavior. Related to this aspect is the ability to perform self-evaluation and reflection: agents may inspect their own outputs, receive feedback from the environment or from other agents, and revise their subsequent actions accordingly [8].

Recent research has also emphasized multi-agent architectures, where complex tasks are distributed among multiple specialized agents. In these systems, agents can assume different roles, such as planner, executor, critic, verifier, domain expert, or coordinator. A manager agent can organize the overall process, assign subtasks to specialized agents, collect their outputs, resolve conflicts, and trigger further refinement when needed [9][10]. Such architectures are especially promising when the target task naturally involves heterogeneous expertise and multiple validation stages.

From this perspective, Agentic AI should not be understood merely as the use of LLMs to generate text, but as an architectural paradigm in which reasoning, planning, tool use, memory, verification, and inter-agent coordination are combined to support complex semi-autonomous workflows. This makes agent-based approaches particularly suitable for the automated or semi-automated generation of structured, multilayer knowledge representations, where the final result depends on the integration of heterogeneous sources, specialized processing modules, and iterative quality control.

III. Challenges in the Production of Rich IEEE 1599 Encodings

Despite the expressive power of IEEE 1599 and the wide range of applications enabled by its multilayer architecture, the creation of richly populated documents remains one of the main obstacles to the large-scale adoption of the format. The issue is not related to the descriptive capabilities of the standard itself, but rather to the effort required to produce and synchronize the heterogeneous information it can host.

A typical encoding workflow begins with the transcription of the original score using a digital score editor such as *MuseScore* [11]. Through dedicated export tools and plugins [12], the symbolic representation can be converted into the *Logic layer* of an IEEE 1599 document. At the same time, the score editor can provide additional resources that are immediately reusable within the IEEE 1599 framework, including graphical score renderings for the *Notational layer* and computer-driven performances that can serve as instances for the *Audio layer*.

The subsequent stages are considerably more demanding. Additional graphical score versions, such as manuscript sources, historical editions, or alternative engravings, must be acquired and linked to the corresponding logical events. This process requires identifying where each musical event appears on the page, typically through bounding boxes, regions, or other graphical coordinates, and linking that area to the corresponding symbolic element.

Similarly, audio and video performances need to be synchronized with the symbolic representation. This process requires the identification of the temporal occurrence of musical events within each recording and the creation of the corresponding mappings toward the spine structure.

Since IEEE 1599 supports multiple instances within the same layer, the effort increases linearly with the number of score versions and performances to be included.

Several support tools have been developed over the years to facilitate these operations. In addition, specific conditions can significantly reduce workload. For example, synchronization can be performed at the measure level rather than at the granularity of individual score symbols, thereby greatly reducing the number of mappings to create. Likewise, recordings generated from computer performances or performances characterized by a highly regular tempo can be aligned more easily than expressive human interpretations presenting substantial agogic variations. Nevertheless, such situations represent favorable cases rather than the norm.

In general, the production of a rich IEEE 1599 document remains a time-consuming activity that requires expert supervision. The effort depends on multiple factors, including the complexity of the notation, the density of musical events, the number of versions of the score and performances involved, and the interpretative variability of the recordings. As an indicative example, the encoding of a relatively small composition consisting of a two-page score, three graphical score versions, and three audio performances may require up to ten hours of manual work. This estimate can vary considerably according to the characteristics of the musical material, but highlights the substantial cost associated with the creation of high-quality multilayer resources.

The challenge of reducing production costs has motivated a significant part of the research conducted around IEEE 1599 after the standardization of the format. Automatic and semi-automatic approaches have been investigated for tasks such as score recognition, audio-to-score synchronization, performance alignment, and feature extraction [13][14]. Although important progress has been achieved, complete automation of the encoding workflow remains an open research problem, particularly when dealing with historical sources, handwritten notation, and expressive performances. Consequently, the availability of large IEEE 1599 collections is still constrained by the amount of human effort required during the creation phase.

These considerations strongly motivate current research into AI techniques, and particularly into agentic AI paradigms, where autonomous software agents can cooperate to perform complex encoding workflows, combining perception, reasoning, synchronization, and validation capabilities within a unified framework.

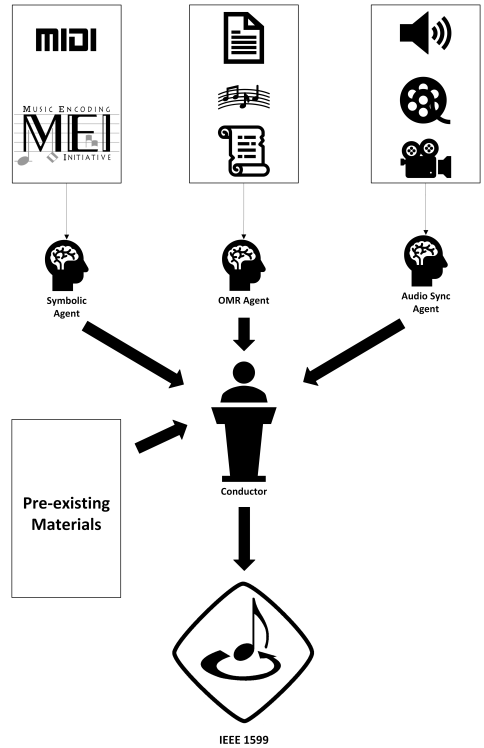

Figure 1. Agent-based workflow for IEEE 1599 document generation.

IV. An Agent-Based Vision for IEEE 1599 Creation

We envision a modular architecture composed of specialized agents, each responsible for generating and/or validating specific portions of an IEEE 1599 document. The proposed framework, depicted in Fig. 1, includes ingestion, transformation, verification, and synchronization stages coordinated by a supervisor agent.

A. Symbolic Format Conversion

The first scenario concerns the automatic transformation of existing symbolic encodings into IEEE 1599. Within the set of available formats, it is worth citing MusicXML [15], and MEI [16], two XML dialects that have established themselves as de facto standards in the field. Other text-based formats, such as Humdrum [17], LilyPond [18], or abc [19], are paired with archives of significant relevance. Many utilities exist for converting between individual formats; while a complete survey is outside the scope of this contribution, it is worth mentioning the Python package *music21* [20].

An AI agent can orchestrate calls to the appropriate tools with the goal of analyzing the source encoding, extracting musical semantics, and generating the corresponding structures in the IEEE 1599 *Logic layer*.

The automation of the conversion task broadens access to the archives and collections encoded with these formats, thus promoting interoperability. In terms of challenges, different formats have different levels of expressivity, so it might not be possible to transfer all the semantics from the original formats. It then becomes paramount to ensure that metadata are preserved and that the original sources are correctly cited, e.g., by using DOIs when available.

B. Optical Music Recognition for the Automatic Population of the Notational Layer

Optical Music Recognition (OMR) refers to techniques aimed at automatically extracting symbolic music notation from images of printed scores, scanned documents, or photographs. Analogously to Optical Character Recognition (OCR) for textual documents, OMR systems identify graphical symbols such as noteheads, rests, clefs, accidentals, articulations, dynamics, and other notational elements, and reconstruct their musical meaning in a machine-readable format.

Within the IEEE 1599 framework, OMR technologies can play a crucial role in the automatic population of the *Notational layer*. Instead of manually encoding a score through a notation editor, recognized symbols can be directly mapped onto the graphical and symbolic structures required by the

 ISBN: 978-1-68558-403-0

format. This capability is particularly relevant in scenarios that involve the digitization of historical archives, the processing of legacy printed scores, and large-scale document-generation workflows.

Several OMR solutions are currently available and can be integrated into web-based or AI-assisted workflows:

- *Audiveris* [21], is the most mature open-source OMR platform. Released under the AGPL license, it recognizes printed Common Western Music Notation and exports the extracted information to MusicXML and to its own XML-based internal format. *Audiveris* combines image processing, rule-based analysis, OCR modules, template matching, and machine-learning classifiers, including neural networks for symbol recognition. Its source code is publicly available and can be deployed locally or integrated into custom processing pipelines;
- *OpenOMR* [22], and related research projects provide open-source implementations of specific OMR functionalities. Although generally less mature than *Audiveris*, they can be useful as building blocks in experimental workflows;
- Commercial services such as *PhotoScore* [23], *SmartScore* [24], *PlayScore* [25], and other cloud-based score-recognition platforms provide proprietary OMR engines. These systems typically require commercial licenses, subscription plans, or dedicated agreements for large-scale integration. Their internal algorithms are generally not publicly documented, limiting their transparency and customizability.

From the perspective of the AI-agent-based generation of IEEE 1599 documents, OMR engines constitute valuable perception modules. An intelligent agent may orchestrate the complete workflow by acquiring score images, invoking an OMR service, validating recognized symbols, and automatically generating the corresponding IEEE 1599 structures. In this context, OMR becomes one of the tools available to the agent for extracting musical knowledge from visual sources.

Audiveris is particularly interesting as a case study because its open architecture allows direct access to recognized musical information prior to export. Consequently, an AI agent could exploit the internal OMR results, MusicXML exports, or confidence measures produced during recognition to populate IEEE 1599 documents while preserving traceability between graphical evidence and symbolic interpretation.

An additional opportunity arises from the layered architecture of IEEE 1599 itself. Beyond the *Notational layer*, the standard includes a *Logic layer*, which encodes the abstract musical content independently of its graphical representation. Traditionally, the *Logic layer* is obtained by manually transcribing a score using notation editors such as *MuseScore*, *Dorico*, or *Sibelius* and subsequently exporting the resulting symbolic representation. However, OMR technologies suggest an alternative workflow in which the *Logic layer* is derived automatically from the symbols recognized in the score image. Such an approach could significantly reduce manual effort and make large-scale digitization projects more feasible.

The relationship between OMR and the *Logic layer* may also operate in the opposite direction. If symbolic information is already available, it can potentially constrain and guide the recognition process. Knowledge of expected pitches, durations, voices, harmonic structures, or formal organization could help disambiguate uncertain graphical interpretations. This concept is analogous to language models supporting OCR through linguistic context. Although contemporary OMR systems generally do not expose interfaces for supplying external symbolic knowledge during recognition, recent research in AI-assisted OMR and multimodal music understanding suggests that such bidirectional interactions may become increasingly relevant in future systems. In this perspective, the *Logic layer* of IEEE 1599 could act not only as a target representation but also as a source of contextual information capable of improving recognition accuracy.

C. Automatic Audio Synchronization for the Population of the Audio Layer

Audio synchronization, often referred to as *score following*, *audio-to-score alignment*, or *music synchronization*, encompasses a family of techniques aimed at establishing temporal correspondences between a symbolic music representation and one or more audio recordings. Given a musical score and a performance, synchronization algorithms estimate the temporal position of musical events within the audio stream, thus creating links between symbolic and acoustic domains.

Within the IEEE 1599 framework, automatic synchronization technologies are particularly relevant to the population of the *Audio layer*. The standard explicitly supports temporal relationships between symbolic music representations and multimedia resources, allowing notes, measures, phrases, and other musical entities (*events*) to be associated with precise temporal locations in audio and video recordings. Traditionally, these relationships are created through manual tempo-tapping activities, which can be labor-intensive and time-consuming. Automatic synchronization provides a scalable alternative by generating temporal mappings directly from the score and the performance.

Several synchronization solutions are currently available and can be integrated into web-based or AI-assisted workflows:

- *MATCH* [26]–[28], and related score-performance alignment tools developed within the Music Information Retrieval (MIR) community provide algorithms for offline synchronization between symbolic scores and audio recordings. Such systems are frequently employed in musicological research and digital music libraries;
- *madmom* [29], is an open-source Python framework for audio and music signal processing. Although primarily designed for tasks such as beat tracking, onset detection, and tempo estimation, its outputs can support synchronization workflows;
- *Commercial music-recognition services*, including proprietary audio-identification and music-analysis platforms, may provide synchronization-related capabilities through application programming interfaces (APIs). However,

these solutions generally require commercial agreements and often expose only high-level functionalities rather than complete alignment data.

From the perspective of AI-agent-based generation of IEEE 1599 documents, synchronization engines can be viewed as temporal interpretation modules. An intelligent agent may automatically acquire a symbolic score, retrieve one or more performances, invoke synchronization algorithms, validate the resulting alignments, and populate the *Audio layer* with the temporal coordinates required by IEEE 1599. This workflow significantly reduces the manual effort traditionally associated with multimedia annotation.

The availability of open-source synchronization libraries is particularly advantageous in this context. Unlike black-box services, open implementations allow AI agents to access intermediate data such as detected note onsets, beat positions, confidence scores, alignment paths, and synchronization errors. These data can be exploited both for automatic validation and for generating provenance information describing how temporal correspondences have been established.

The layered architecture of IEEE 1599 offers additional opportunities for integrating synchronization technologies. In particular, the *Logic layer* provides a symbolic representation of musical events that can serve as the basis for alignment. In this way, algorithms may operate on pitches, durations, measures, voices, and structural information extracted from the *Logic layer*. This approach decouples synchronization from a specific graphical representation and enables the reuse of symbolic knowledge across multiple performances. More generally, the interaction between symbolic representations and temporal observations suggests a bidirectional relationship in which the *Logic layer* supports synchronization while synchronization provides empirical evidence about the realization of the symbolic score.

Synchronization results can also enrich the *Structural layer*. Repeated alignments across multiple performances may reveal expressive timing patterns, performance conventions, or structural ambiguities that can be incorporated into higher-level analytical descriptions [30].

Recent advances in AI further strengthen this perspective. Modern synchronization systems based on deep learning increasingly exploit multimodal representations that combine symbolic, graphical, and acoustic information. Consequently, AI agents can orchestrate heterogeneous synchronization resources, select the most appropriate alignment strategy for a given repertoire, compare alternative results, and automatically generate IEEE 1599 documents containing symbolic and temporal information. In such workflows, synchronization becomes a key enabling technology for bridging the gap between musical notation and musical performance.

V. Open Problems

Although promising, the use of these tools is not without challenges. In this section, we introduce, without necessarily claiming solutions, some of the most significant ones encountered in the conceptualization of this idea.

A. Reliability and Validation

When it comes to LLMs, users are nowadays acquainted with the so called “hallucination” phenomena, the generation of “plausible yet not factual content” [31]. Errors and hallucinations should be detected and corrected early in order to prevent propagation from one task to the next. In this sense, the adoption of the Human-in-the-Loop approach, a strategy where users step in to supervise, approve, edit, or reject the output of a task, seems clearly advantageous.

B. Interoperability, Explainability and Provenance

In the current AI cloud infrastructure, there are many providers with their different models. It would be interesting to test if different models have vastly different outcomes for specific tasks. In this case, it would become relevant to determine how the orchestration between heterogeneous tools can be ensured.

A further topic of interest is AI explainability and interpretability [32], which refers to methods and processes that allow users to understand, and ultimately trust, the processes by which the model generated a result.

To enhance transparency and trust in AI-generated content, it is essential to maintain a verifiable record of the data’s provenance and revision history. Although IEEE 1599 does not natively support this functionality, repository-based workflows and versioning systems, such as Git, can complement the format by documenting the origin, evolution, and authorship of the encoded content.

C. Ethical considerations

Especially when using commercially available tools, it is worth asking how these models have been trained. While the legal aspects are still being adjudicated in courts, the presence of copyrighted materials in the training data is to be assumed unless the developers explicitly state otherwise.

Similarly, when interacting with these systems, the question of what happens to shared or uploaded data becomes critical. Do such systems collect and retain user-provided data, and might these data subsequently be used for further training purposes? One possible solution is the development of custom, on-premises agents that are not connected to the internet and can be trained or fine-tuned exclusively on materials for which the relevant copyrights are owned or properly licensed. Although this approach may limit their applicability to general-purpose tasks, it may also increase their reliability, security, and domain-specific accuracy.

A further issue to consider is the resource demand of these models. The proliferation of dedicated data centers, with their associated land use and electricity and water consumption, deserves careful scrutiny.

VI. Conclusions

AI-based agents have the potential to significantly simplify the creation of IEEE 1599 documents by automating symbolic conversion, score recognition, and audio synchronization tasks. While several technical and methodological challenges remain open, the integration of agentic AI into music representation workflows may contribute to the development of next-generation music archives and intelligent cultural heritage systems.

Future work should investigate prototype implementations and evaluation methodologies for agent-based IEEE 1599 generation pipelines.

ACKNOWLEDGMENTS

This project is partially supported by KreativInstitut.OWL, a consortium consisting of OWL University of Applied Sciences and Arts, Detmold University of Music, and Paderborn University, funded by the Ministry of Economic Affairs, Industry, Climate Action and Energy of the State of North Rhine-Westphalia, Germany.

REFERENCES

[1] G. Haus and M. Longari, "A multi-layered, time-based music description approach based on XML," *Computer Music Journal*, vol. 29, no. 1, pp. 70–85, 2005.

[2] D. L. Baggi and G. M. Haus, *Music navigation with symbols and layers: Toward content browsing with IEEE 1599 XML encoding*. John Wiley & Sons, 2013.

[3] L. Wang et al., "A survey on large language model based autonomous agents," *Frontiers of Computer Science*, vol. 18, no. 6, p. 186 345, 2024.

[4] Z. Xi et al., *The Rise and Potential of Large Language Model Based Agents: A Survey*, 2023. arXiv: 2309.07864 [cs.AI].

[5] X. Huang et al., "Understanding the Planning of LLM Agents: A Survey," *arXiv preprint arXiv:2402.02716*, 2024. [Online]. Available: https://arxiv.org/abs/2402.02716.

[6] J. S. Park et al., "Generative Agents: Interactive Simulacra of Human Behavior," in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '23, San Francisco, CA, USA: Association for Computing Machinery, 2023, ISBN: 9798400701320. DOI: 10.1145/3586183.3606763.

[7] Z. Zhang et al., "A Survey on the Memory Mechanism of Large Language Model based Agents," *arXiv preprint arXiv:2404.13501*, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2404.13501.

[8] N. Shinn et al., "Reflexion: Language Agents with Verbal Reinforcement Learning," in *Advances in Neural Information Processing Systems*, 2023.

[9] Q. Wu et al., *AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation*, 2023. arXiv: 2308.08155 [cs.AI].

[10] X. Li, S. Wang, S. Zeng, Y. Wu, and Y. Yang, "A survey on LLM-based multi-agent systems: workflow, infrastructure, and challenges," *Vicinagearth*, vol. 1, no. 1, p. 9, Oct. 8, 2024, ISSN: 3005-060X. DOI: 10.1007/s44336-024-00009-2.

[11] *MuseScore*, https://musescore.org/, [Online; accessed 24-June-2026].

[12] *IEEE 1599 MuseScore Plugin*, https://github.com/LIMUNIMI/musescorePlugins, [Online; accessed 24-June-2026].

[13] S. Baldan, L. A. Ludovico, and D. A. Mauro, "Algorithms for an Automatic Transcription of Live Music Performances into Symbolic Format," in *Auditory Display. Genesis of Meaning in Sound and Music. 6th International Symposium, CMMR/ICAD 2009, Copenhagen, Denmark, May 18-22, 2009 : revised papers*, ser. Lecture Notes in Computer Science, vol. 5954, Berlin, Heidelberg: Springer-Verlag, 2010, pp. 422–437, ISBN: 978-3-642-12438-9 (print), 978-3-642-12439-6 (online). DOI: 10.1007/978-3-642-12439-6_22.

[14] A. Baratè et al., "Web Applications for Automatic Audio-to-Score Synchronization with Iterative Refinement," in *Proceedings of the Sound and Music Computing Conference 2023, Stockholm, Sweden*, ser. Proceedings of the Sound and Music Computing Conferences, SMC Network, 2023, pp. 154–161, ISBN: 978-91-527-7372-7.

[15] *MusicXML*, https://www.musicxml.com/, [Online; accessed 24-June-2026].

[16] *MEI*, https://music-encoding.org, [Online; accessed 24-June-2026].

[17] *Humdrum*, https://www.humdrum.org/, [Online; accessed 24-June-2026].

[18] *LilyPond*, https://lilypond.org/, [Online; accessed 24-June-2026].

[19] *abc notation*, https://abcnotation.com/, [Online; accessed 24-June-2026].

[20] *music21*, https://music21.org/, [Online; accessed 24-June-2026].

[21] *Audiveris*, https://audiveris.com/, [Online; accessed 24-June-2026].

[22] *OpenOMR*, https://sourceforge.net/projects/openomr/, [Online; accessed 24-June-2026].

[23] *PhotoScore*, https://www.neuratron.com/photoscore.htm, [Online; accessed 24-June-2026].

[24] *SmartScore*, https://www.musitek.com/, [Online; accessed 24-June-2026].

[25] *PlayScore*, https://www.playscore.co/, [Online; accessed 24-June-2026].

[26] S. Dixon and G. Widmer, "MATCH: A Music Alignment Tool Chest.," in *Proceedings of the 6th International Conference on Music Information Retrieval*, (London, United Kingdom), ISMIR, Sep. 2005, pp. 492–497. DOI: 10.5281/zenodo.1416952.

[27] S. D. Peter et al., "Automatic Note-Level Score-to-Performance Alignments in the ASAP Dataset," *Transactions of the International Society for Music Information Retrieval*, Jun. 2023. DOI: 10.5334/tismir.149.

[28] J. Zeitler, B. Maman, and M. Müller, "Robust and Accurate Audio Synchronization Using Raw Features From Transcription Models," in *Proceedings of the 25th International Society for Music Information Retrieval Conference*, (San Francisco, California, USA and Online), ISMIR, Nov. 2024, pp. 120–127. DOI: 10.5281/zenodo.14877000.

[29] S. Böck, F. Korzeniowski, J. Schlüter, F. Krebs, and G. Widmer, "Madmom: A new python audio and music signal processing library," in *Proceedings of the 24th ACM international conference on Multimedia*, 2016, pp. 1174–1178.

[30] A. Baratè, G. Haus, L. A. Ludovico, and G. Presti, "Investigating Interpretative Models in Music through Multi-Layer Representation Formats," *Journal of Music, Technology & Education*, vol. 12, no. 1, pp. 95–111, 2019, ISSN: 1752-7066 (print), 1752-7074 (online). DOI: 10.1386/jmte.12.1.95_1.

[31] L. Huang et al., "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions," *ACM Trans. Inf. Syst.*, vol. 43, no. 2, Jan. 2025, ISSN: 1046-8188. DOI: 10.1145/3703155.

[32] ISO/IEC, *Software and systems engineering — software testing — part 11: Guidelines on the testing of ai-based systems*, Standard, Geneva, Switzerland, 2020. [Online]. Available: https://www.iso.org/standard/79016.html.