Musical Agent Systems: MACAT and MACataRT

Keon Ju M. Lee*
School of Interactive Arts and Technology
Simon Fraser University
Vancouver, British Columbia, Canada
keon_maverick@sfu.ca

Philippe Pasquier*
School of Interactive Arts and Technology
Simon Fraser University
Vancouver, British Columbia, Canada
pasquier@sfu.ca

Abstract

Our research explores the development and application of musical agents—human-in-the-loop generative AI systems designed to support music performance and improvisation within co-creative spaces. We introduce MACAT and MACataRT, two distinct musical agent systems crafted to enhance interactive music-making between human musicians and AI. MACAT is optimized for agent-led performance, employing real-time synthesis and self-listening to shape its output autonomously, while MACataRT provides a flexible environment for collaborative improvisation through audio mosaicing and sequence-based learning. Both systems emphasize training on personalized, small datasets, fostering ethical and transparent AI engagement that respects artistic integrity. This research highlights how interactive, artist-centred generative AI can expand creative possibilities, empowering musicians to explore new forms of artistic expression in real-time, performance-driven and music improvisation contexts.

1 Introduction

1.1 Background

A **musical agent** [1] is an artificial system designed to automate creative musical tasks within the field of **musical metacreation**—an area focused on the computational simulation of musical creativity [2]. In this domain, musical agents perform a variety of crucial tasks that deepen and broaden human-AI collaboration in music-making. These tasks encompass generative composition, where agents autonomously produce original material based on learned styles; interactive performance, which allows agents to adapt in real-time to live inputs from human musicians; and accompaniment, in which agents provide dynamic, context-sensitive support that complements primary musical performances. Further, style adaptation enables agents to align their outputs with specific genres or artist preferences. At the same time, real-time improvisation allows for spontaneous, unscripted musical responses, fostering a creative and responsive interaction with human musicians. These tasks position musical agents as versatile, adaptive collaborators, functioning as co-creators, responsive partners, and innovative contributors across diverse musical contexts.

Drawing on principles from Artificial Intelligence (AI [3]) and Multi-Agent Systems (MAS [4]), musical agents operate autonomously, making decisions and performing actions in response to real-time musical contexts. These systems range from simple rule-based models to advanced frameworks capable of learning and adapting through interaction. Often implemented using MAX/MSP programming [5], they facilitate collaboration with human musicians or other agents, showcasing attributes such as reactivity, adaptability, and coordination. Complementing these agents, Corpus-Based Concatenative Synthesis (CBCS)—pioneered in IRCAM's CataRT system [6] and widely used in electroacoustic music [7,8,9]—utilizes small, personalized datasets of recordings to generate

*Metacreation Lab for Creative AI: https://www.metacreation.net/

Workshop on Creativity & Generative AI (NeurIPS 2024).

new content. CBCS achieves this by selecting and transforming audio segments from a pre-existing corpus, enabling real-time control of the generative process while preserving the expressive nuances of human performance.

1.2 Research Motivation

Our research is motivated by the aim of developing interactive, artist-in-the-loop generative AI systems designed for musicians and sound artists. The primary objective is to enable musicians to explore novel creative possibilities and expand their musical and artistic practices, particularly within real-time scenarios and interactive environments.

Our musical agent systems, MACAT and MACataRT, are based on the mindset of using small data in music [10] and the principles of model crafting in visual arts with generative AI [11], as well as fostering music-making in co-creative spaces between human musicians and musical agents [12], including machine listening [13].

Training on a small, high-quality audio corpus enables our musical agent systems to closely align with the specific musical nuances and stylistic preferences of collaborating musicians. In contrast to models that broadly reproduce generalized patterns derived from large-scale audio or MIDI datasets, our model is optimized to interpret and respond to the unique attributes embedded within a musician's work, thereby facilitating the personalization of music systems. This focused approach enhances the agent's capacity to function as a genuine creative collaborator, fostering stylistic coherence and adaptability that could otherwise be diluted in models trained on more generic, large-scale audio or MIDI corpora.

Our musical agent systems are highly flexible, enabling artists to train the models using curated datasets tailored specifically to their unique stylistic preferences. This approach, which we term "model crafting," allows artists to incorporate their own recorded data into the training process, creating an agent that is not bound by predefined genres or styles but is instead highly personalized to their artistic identity. By selecting and shaping the training corpus, artists can ensure that the musical agent aligns closely with their aesthetic vision, resulting in outputs reflecting their individual voices and creative nuances. This flexibility makes the system adaptable to a wide range of musical expressions, empowering artists to use AI not as a generic accompaniment tool but as a customizable co-creator that can evolve with their artistic practice.

2 Workflow of Musical Agent Systems

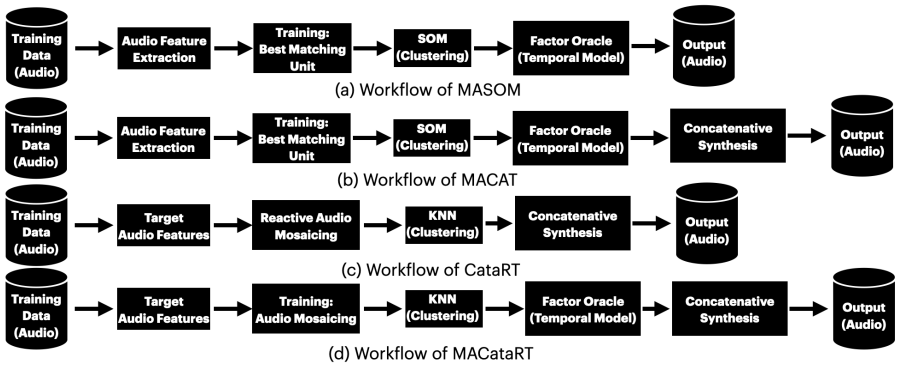

Figure 1: The workflow of each musical agent system for the comparison: (a) MASOM, (b) MACAT, (c) CataRT, and (d) MACataRT.

2.1 Workflow of MASOM and MACAT

Musical Agent based on Self-Organizing Maps (MASOM [14]) is a machine improvisation system for live performances, particularly suited for experimental music and free improvisation. It integrates self-organizing maps (SOM [15]) as a sound memory, Variable Markov Models (VMM [16]) to recognize and generate musical structures, and affective computing for real-time audio analysis.

MASOM can listen to and generate audio signals, learning from a corpus of recordings to structure its performances. Appendix B provides a comprehensive explanation of the latest version of MASOM.

MACAT, as shown in the workflow in Figure 1 (b), is an enhanced version of the MASOM agent developed by the Metacreation Lab, featuring integrated real-time sound synthesis and visualization. At its core, MACAT utilizes concatenative sound synthesis and offers improved visualization for past and current nodes. It enables sound artists to flexibly synthesize audio segments and explore a broad range of timbral possibilities through sound design parameters. It employs the SOM and Factor Oracle (FO [17]), a suffix automaton for real-time pattern recognition in sequences of nodes representing clusters of audio segments grouped by timbral similarity. During training, MACAT conducts offline machine listening to analyze the original audio data, initialize SOM nodes, and identify the Best Matching Unit (BMU) for each input vector based on Euclidean distance. The BMU and its neighbours are adjusted, resulting in a SOM that clusters audio segments into groups. Subsequently, MACAT learns the sequence of nodes using a VMM, with the FO identifying patterns within this sequence during real-time generation. Dynamically adapting to its self-listening process, MACAT uses a congruence parameter to regulate FO's forward and backward transitions, ensuring responsive and context-sensitive generative output for live performances and creative exploration.

2.2 Workflow of CataRT and MACataRT

The IRCAM's CataRT system [6] is a real-time corpus-based concatenative synthesis tool designed for interactive sound exploration by selecting sound units from a database based on audio descriptors. Implemented in Max/MSP, the system organizes sounds within an audio descriptor space, enabling users to target specific audio features, such as pitch or timbre, as illustrated in the workflow of Figure 1 (c). Building on traditional granular synthesis, the system utilizes a curated corpus of segmented audio crafted by artists, enabling precise control over sound characteristics through concatenative sound synthesis. Its versatile applications span interactive sound synthesis, gesture-controlled synthesis, live audio resynthesis, and expressive speech synthesis, offering a flexible and powerful interface for creative sound exploration.

MACataRT, as depicted in the workflow in Figure 1 (d), is an enhanced version of the CataRT system, incorporating a temporal model based on the factor oracle and offering a more intuitive interface for sound synthesis and resampling. The original CataRT system, while capable of reactive improvisation through real-time machine listening, clusters audio segments using K-Nearest Neighbors (KNN [18]) and targets specific audio features for audio mosaicing. However, it lacks a temporal model to manage time-based musical structures. To address this limitation, MACataRT integrates the factor oracle to automate the generation process, building on CataRT's audio mosaicing capability. Audio mosaicing [19] assembles audio segments into new audio pieces by selecting segments that match target audio features specified by musicians, who then refine the output to replicate the desired characteristics. MACataRT enhances this process with interactive audio mosaicing that functions in both real-time and offline modes. In real-time, the musical agent facilitates reactive improvisation, responding to live inputs based on machine listening and targeting audio features without using the factor oracle. In its proactive improvisation mode, the musical agent system learns sequences of audio segment indices during offline training, enabling the factor oracle to generate music based on these learned sequences. This dual capability allows musicians to interactively play alongside MACataRT, fostering dynamic, creative exchanges. Details on the interface of MACAT and MACataRT systems are provided in Appendix C.

3 Research-Creation and Musical Practice

3.1 Research-Creation Methodology

The research-creation methodology in musical practice [20], especially in performances with musical AI agents, combines scholarly inquiry with creative experimentation, focusing on co-creation between humans and AI systems. Musicians interact with AI agents as collaborative partners, capable of responding to real-time inputs and aligning with the performer's expressive intent. This relationship fosters a dynamic musical dialogue where AI agents generate, adapt, and influence live performances, broadening the possibilities for improvisation, stylistic adaptation, and spontaneous composition. This approach merges artistic exploration with technical research, offering insights into AI's creative potential in music and redefining the musician's role in AI-augmented performances.

The research-creation methodology is essential in musical performance and improvisation, enabling a deeper and more nuanced exploration of the experiential and artistic aspects of music-making that quantitative analysis alone cannot fully capture. In such contexts, emphasis is placed on interpretive, spontaneous, and affective qualities that emerge through real-time collaboration, particularly with musical AI agents. While quantitative analysis offers insights into measurable elements like timing accuracy and frequency distribution, it fails to address the subjective, responsive, and context-sensitive dimensions of co-creation that are fundamental to live musical practice. Musical agent systems, regarded as intelligent musical instruments, require research-creation to cultivate these interactive relationships, allowing artists to develop virtuosity in collaboration with AI, create aesthetically compelling music, and explore new or unique musical styles. Given the absence of well-established quantitative measures for evaluating personalized improvisation systems and diverse musical scenarios, our approach prioritizes showcasing musical practice and performance over suboptimal computational and quantitative evaluation.

3.2 Musical Application and Artistic Practice

The application of each musical agent varies, offering musical affordances that support the research-creation in real-world musical performance scenarios. MACAT, for instance, is particularly effective in solo performance settings, where its improvisational output adapts through a self-listening process that allows it to take the lead in most musical contexts. MACAT was showcased at the MusicAcoustica Festival in Hangzhou, China, by the artist collective K-Phi-A, featuring Philippe Pasquier on live ambient electronics, Keon Ju Maverick Lee on percussions, and VJ Amagi providing audiovisuals.

MACataRT expands creative possibilities through audio mosaicing, supporting both reactive and proactive improvisation and proving its adaptability across diverse collaborative settings. Its practical effectiveness was demonstrated in a live performance, where it was used by a percussionist (Keon) and a guitarist (Sara) to co-create music. This collaboration earned significant recognition when their piece, *Echoes of Synthetic Forest* by the music duo *KeRa*, was selected as one of the Top 10 finalists in the 2024 AI Music Song Contest [21] and performed in Zürich, Switzerland [22]. These achievements highlight the value of research-creation, showcasing how musical agents can enhance real-time artistic expression and foster collaborative exploration. Our musical agent systems and showcases are available for public access on the Metacreation Lab's GitHub repository[2].

4 Conclusion and Future Work

Our research highlights the potential of the musical agents MACAT and MACataRT to enable creative collaboration between AI and human musicians, offering a novel approach to interactive music-making through real-time generative AI. These systems prioritize the preservation of expressive nuances by employing corpus-based concatenative synthesis and small-data training methods, allowing the musical agents to act as responsive co-creators in diverse performance contexts. MACAT and MACataRT demonstrate how artist-in-the-loop AI agents can significantly broaden the creative options for musicians, providing practical tools that integrate into live performances and improvisational settings. Ethical considerations related to our musical agents are discussed in Appendix A.

In future work, we aim to enhance both the temporality and explainability of our musical agent systems. Currently, these systems utilize conventional audio mosaicing techniques and the factor oracle algorithm for temporal modelling to identify and generate musical patterns. To improve temporality, we plan to integrate deep learning architectures that enable agents to learn and retain longer sequences of musical patterns. For enhanced explainability, we intend to incorporate a module that records the history of past musical patterns, thereby advancing comprehension from a note-level to a bar-level understanding. Additionally, we aim to advance the machine listening module to deepen the agents' musical comprehension. A feedback loop using reinforcement learning may also be introduced to further enhance the adaptability of musical agents in real-time performance contexts.

[2] https://github.com/Metacreation-Lab/Musical-Agent-Systems

References

[1] Tatar, K., & Pasquier, P. (2019). Musical agents: A typology and state of the art towards musical metacreation. Journal of New Music Research.

[2] Pasquier, P., Eigenfeldt, A., Bown, O., & Dubnov, S. (2017). An introduction to musical metacreation. Computers in Entertainment (CIE)

[3] Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach. Pearson.

[4] Wooldridge, M. (2009). An introduction to multiagent systems. John Wiley & sons.

[5] Puckette, M. (1991). Combining event and signal processing in the MAX graphical programming environment. Computer music journal.

[6] Schwarz, D., Beller, G., Verbrugghe, B., & Britton, S. (2006). Real-time corpus-based concatenative synthesis with CataRT. In the proceedings of the Conference on Digital Audio Effects (DAFx).

[7] Sturm, B. L. (2006). Adaptive Concatenative Sound Synthesis and Its Application to Micromontage Compositior. Computer Music Journal.

[8] Hackbarth, B., Schnell, N., Esling, P., & Schwarz, D. (2013). Composing morphology: Concatenative synthesis as an intuitive medium for prescribing sound in time. Contemporary Music Review.

[9] Einbond, A., Bresson, J., Schwarz, D., & Carpentier, T. (2021). Instrumental Radiation Patterns as Models for Corpus-Based Spatial Sound Synthesis: Cosmologies for Piano and 3D Electronics. In the proceedings of the International Computer Music Conference (ICMC).

[10] Vigliensoni, G., Perry, P., & Fiebrink, R. (2022). A small-data mindset for generative AI creative work. In the proceedings of the Machine Learning for Creativity and Design Workshop at the conference on Neural Information Processing Systems (NIPS).

[11] Abuzuraiq, A. M., & Pasquier, P. (2024). Seizing the Means of Production: Exploring the Landscape of Crafting, Adapting and Navigating Generative AI Models in the Visual Arts. In the proceedings of the Computer-Human Interaction (CHI) workshop on Generative AI and HCI.

[12] Thelle, N. J. W., & Wærstad, B. I. G. (2023). Co-Creatives Spaces: The machine as a collaborator. In Proceedings of the Conference on New Interfaces for Musical Expression (NIME).

[13] Rowe, R. (1992). Interactive music systems: machine listening and composing. MIT press.

[14] Tatar, K. & Pasquier, P. (2017). MASOM: A Musical Agent Architecture based on Self-Organizing Maps, Affective Computing, and Variable Markov Models. In the proceedings of the Conference on International Workshop on Musical Metacreation (MuMe).

[15] Kohonen, T. (1990). The self-organizing map. In the proceedings of the conference on the Institute of Electrical and Electronics Engineers (IEEE).

[16] Begleiter, R., El-Yaniv, R., & Yona, G. (2004). On prediction using variable order Markov models. Journal of Artificial Intelligence Research

[17] Assayag, G., & Dubnov, S. (2004). Using factor oracles for machine improvisation. Journal of Soft Computing.

[18] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE transactions on information theory.

[19] Lazier, A., & Cook, P. (2003). MOSIEVIUS: Feature driven interactive audio mosaicing. In the proceedings of the conference on Digital audio effects (DAFx).

[20]Stévance, S., & Lacasse, S. (2017). creation in Music and the Arts: Towards a Collaborative Interdiscipline. Routledge.

[21] Top 10 Finalist AI Music Song Contest (2024)
URL: https://www.aisongcontest.com/the-2024-finalists (Last accessed: 14th of November, 2024).

[22] KeRa's Music Performance as a Top 10 Finalist at Zürich, Switzerland (2024)
URL: https://www.youtube.com/live/2ojQEGNUXic?si=nEmzbR_yCyhyk1aS&t=4499 (Last accessed: 14th of November, 2024).