Integrating Signal Processing and Computational Creativity: A Theoretical Framework for Modern Music Composition

Andrew Zhengyu Hu

Tsinghua International School, Beijing, China
andrewmingren@outlook.com

Abstract. The field of music composition is undergoing a significant transformation due to advancements in artificial intelligence. This paper explores a theoretical framework for modern music composition by integrating signal processing techniques with computational creativity. The author has sought to analyze how AI, particularly through the application of Differentiable Digital Signal Processing (DDSP) and neural networks, can interpret and generate music based on its structural principles. The discussion includes the use of Markov models to describe music structure as probability patterns, enabling AI to create new pieces while adhering to established principles. Furthermore, the author examines the incorporation of brain-inspired architectures in neural networks to mimic cognitive processes such as key and mode recognition, fostering a synergistic relationship between human composers and AI. This paper also addresses the challenges faced in AI music composition, including long-term structure modeling, dataset bias, and the lack of 'humanness' in AI-generated music. Finally, we discuss the potential for human-AI collaborative composition, emphasizing the evolving role of AI as a partner rather than a replacement for human creativity. This research highlights the ongoing progress and future directions in leveraging AI to enhance and redefine the music composition process.

Keywords: DDSP, End-to-End Music Composition, AI-generated Music

1. Introduction

Nowadays, the field of music composition, like other arts, are deeply influenced by technological developments. With it's rapid advancement, artificial intelligence, which has become a prominent factor that is reshaping many industries, will no doubt affect music composition as well. The recent release of the generative artificial intelligence SUNO allows people to create a piece by simply entering a sentence prompt; similar AIs like OpenAI's transformer-based MuseNet and Aiva has also gained attraction over time. Google's Nsynth allows remixing different instruments for nuanced blends for mixing and editing. While questions remain around how people will evaluate AI generated music, it by no means undermines the potential of artificial intelligence. In this paper, we will analyze the potential of integrating artificial intelligence into the process of music composition.

One of the key aspects of this integration is the ability of AI to interpret and generate music based on its structural principles. Music, like language, has its own set of grammar and structure, which shapes how humans create and perceive it. Unlike language, which relies on natural language

processing, music requires a different method of description. Markov models, for instance, describe music structure as probability patterns, allowing AI to generate music based on these patterns [1]. By combining Markov models with symbolic AI, we can create music that adheres to structural principles while incorporating randomness to generate new pieces [2]. The Markov models describe the music structure as probability patterns. Combining it with symbolic AI, the model can create music using a set of principles while also using randomness to create new pieces. Furthermore, by incorporating brain-inspired architecture in the neural network, we can mimic processes common to music composition such as key and mode recognition, allowing the human composer and artificial intelligence to work in synergy. Neural networks like DDSP can also be used to implement signal processing operations. In the past, components like oscillators, filters, and envelopers must be manually configured. Now, this process can be incorporated within artificial intelligence. By internalizing the whole music making process, we provide detailed information about entire pieces of music to AI. We can use gradient optimization to tune the specific elements of the music. Using backpropagation, AI changes each parameter of the music until the desired output. While discussing the viability of artificial intelligence's contribution to music composition, we must also understand its perception. An aspect of perception is formed by it's difference to the norm, or human made music. As AI's engagement with music is relatively new, it will have a noticeable difference to music composed by humans, which can be found through comparison and evaluation. We will dissect the music based on a set of metrics. By comparing AI generated music and human made music through these metrics, existing literatures found a series of commonalities within AI generated music that opposes the man-made ones. The incorporation of brain-inspired architectures in neural networks enables AI to mimic cognitive processes common to music composition, such as key and mode recognition. This allows for a synergistic relationship between human composers and AI, where the strengths of each can be leveraged to create music that is both innovative and emotionally resonant [3]. For example, spiking neural networks inspired by neuroscience and psychology have been proposed to learn musical modes and keys through evolutionary neural circuits, bridging the gap between cognitive theories and AI-driven composition [3].

In reviewing relevant literatures, another significant advancement in this field has been emphasized: the integration of traditional signal processing techniques with neural networks. Differentiable Digital Signal Processing (DDSP) has emerged as a paradigm for combining traditional signal processing methods, such as additive synthesis and filters, with neural networks. This integration enables controllable music generation, allowing further research to embark on for detailed manipulation of sound parameters. As AI-generated music becomes more prevalent, it is crucial to evaluate its impact and reception. One aspect of this evaluation is the perception of AI-generated music in comparison to human-made music. Since AI's engagement with music is relatively new, there will likely be noticeable differences between AI-generated and human-composed music. These differences can be explored through a set of metrics, including creativity, structural complexity, and emotional impact [4]. By dissecting AI-generated music based on these metrics, we can identify trends and commonalities within AI-generated music and understand how it differs from human-made music.

2. Literature review

With the vast number of sources online, it's expected that many similar sources will show up during the research and not all of them will provide the desired information. For example, while multiple sources may fit under simple keywords like "The role of Differentiable Digital Signal Processing", they aren't always on topic: One research paper, "MIDI-DDSP: Detailed Control of Musical

Performance via Hierarchical Modeling", is about the Using DDSP to extract performance quality. While it is related to DDSP and Music, it strays from the focus of this research. Thus, filtering based on purpose can narrow the field and draw out relevant sources. Another factor used to preview a source's relevancy is its publishing date. To keep the information synthesized in this research up to date, the sources used must be reflective of modern innovation.

Reading the chosen papers, gave insight to the effects of AI in music composition. DDSP plays a large role in the composing process. By integrating classic digital processing elements with neural networks, we can achieve more data-efficient synthesis along with the fine-tuning over individual variables [5]. It enables a hierarchical model that gives user detailed controls while maintaining realistic audio [6]. Using DDSP, models can be trained to alter timbre as a metric; timbre is described as a music's tone quality. Real-timbe timbre has been achieved by Franco Caspe [7]. End-to-End Music Composition is a method in which the music is fully processed from start to end. The symbolic nature allows people to have better control over stylistic aspects of music. As the process of training AI involves many sources of data, problems related to its capability and credibility arise. While AI has made significant improvements in making Human-like music, it still struggles to create long and strong melodies [4]. Aspects that make art unique and human proved to be a challenge for AI; its lack of perspective and worldview means that it struggles to generate rich music [8]. Further research would be needed to identify their genre-specific capabilities [9]. The copyright behind AI is still blurred as different companies uphold varying standards. Nevertheless, for composers to collaborative with AI effectively, they both needs to look at music from the same perspective. Furthermore, conditions may vary with each composer, adding to the complexity of circumstances [10].

Diving into literature concerning Artificial Intelligence's role in music composition, it's apparent that new modes of artificial intelligence specified for music interpretation is required for a collaborative experience between the AI and the composer. DDSP has made significant progress in the field, allowing the most control over different aspects of the music while maintaining the realistic tune. End-to-End Music composition is the preferred way to incorporate entire audio into the processing, making the generation process more precise and efficient. Nonetheless, AI still struggles in other aspects. Its output worsens as more complex subjects are requested, devoid of human emotions associated with music. Experts question whether its generation can be original or copyrighted. In the following chapters, I will explain each of the progress made towards a coherent collaboration between humans and AI, and roadblocks we currently face.

3. The role of DDSP in music composition

Neural models had already been used to tackle music composition, but they still face glaring issues [5]. Neural synthesis generally suffers from a lack of bias because they generate waveforms directly. Time domain models struggle with precise alignment [5]. Fourier-based models face issues with spectral leaks and phase alignment, blurring the audio [5]. Autoregressive models prove to be more accurate but require a larger set of training data [5].

Differentiable Digital Signal Processing, DDSP, can achieve generations with higher fidelity [5]. Its full differentiability allows automatic tuning of parameters that otherwise would require manual control [5]. Engel et al. [5] has shown the result synthesized using DDSP accurately mimics their violin dataset. The model is capable of extrapolating new conditions that were not part of the training dataset, such as shifting to new octaves [5]. Furthermore, DDSP addresses the challenges faced in Singing voice synthesis – pitch, timing accuracy – through differentiable oscillators that adjusted parameters using backpropagation [11].

Caspe et al.'s [7] implementation of DDSP as a real-time virtual synthesizer plugin focuses on timbre transfer and MIDI control. It uses an end-to-end process based on Spectral Modeling Synthesis and an autoencoder neural network architecture to analysis features and predict parameters, allowing the process of timbre transfer [7]. Further research is needed to improve the sound quality of real-time models [7].

3.1. End-to-end music composing

End-to-end music composition, a fully internalized generation process, combines multiple stages of music composition into one streamlined processed that requires minimal manual input. Previously, most generation uses symbolic representations – MIDI - rather than audio-based approaches [9]. System targeted to produce polyphonic music but are often restricted to specific music genres [12].

We can observe many implementations of end-to-end models. Auto NLMC is an encoder-decoder sequential recurrent neural network; its end-to-end nature helps minimize the total loss function [13]. Using lyrics2vectors, dense representation of lyrics facilitates the model's learning of lyrics-melodies relationship; it's capable of producing original lyrics with corresponding melodies [13]. It can be improved by interjecting prior musical knowledge to the model [13].

Similarly, Mao et al.'s [12] Biaxl LSTM is an end-to-end generative model, turning raw MIDI data into complete musical pieces. Its end-to-end nature allows the model to learn to incorporate styles directly through the data. Adaptive temperature adjustments can improve the quality of audio without manual intervention.

MuseNet undergoes an almost completely end-to-end composition process. Its autonomous generation creates complete musical structures based on the user's input, but the final output requires extra manual processing to become audio. Ferreira et al. [14] emphasizes future training on larger network models should explore pretrained embeddings.

3.2. Challenges in AI music composition

Over time, AI's involvement in music saw growing interest, leading to competitions [15]. Searching algorithms and recommendation systems revolving around music reflect similar interests in analyzing music through neural networks [16]. Despite the progress made in AI music composition, the field still faces aesthetic and ethical challenges.

They struggle with long-term structure modeling, interconnected axes of pitches and time [15]. Dataset bias becomes especially relevant with the already limited dataset in this field [15].

The product of artificial intelligence still lacks a sense of "humanness" compared to that composed by people. They still lack high-quality melodies compared to those composed by humans [4]. Rohrmeier [8] describes the embodiment challenge, suggesting that artificial intelligence struggles to consider the human-body experience, the instruments available and performative settings. Rohrmeier [8] tackles a similar problem by evaluating human creativity: creativity should focus on the product rather than the intent. Its relative nature means that artificial intelligence is inherently capable of being creative.

Copyright is complex when it comes to content "created" by artificial intelligence. An existing study suggests that understanding if the work is protected by EU copyright laws requires a distinction of whether this work truly reflects the creator: whether the work reflects human effort and creative choices [17]. Bulayenko et al. [18] believes that further research is needed to determine the practicality of introducing new laws.

3.3. Human-AI collaborative composition

The effectiveness of AI in music composition should be evaluated with consideration of the specific role they play in the process. Co creation is the go-to process where composers work with artificial intelligence as partners: both composers interact creatively on a single goal [19]. While AI is capable of collaborating with people, many people still viewed AI as a tool [19]. Current AI, capable of composition, arrangement, lyric writing, and mixing, can collaborate with novice composers who have little to no experience in production [19].

This shifts the creative process, creating a new stage in the production where composers combine AI generated output with human-composed elements [19]. Though Fu et al. [19] also suggests that the lack of experience may affect composer's collaboration: their lack of knowledge will affect the choice of generation.

The effectiveness of collaboration, according to Gianet et al. [10], depends on factors such as personal motivation, artistic sensibilities and the broader social and cultural context.

Ultimately, Gianet et al. [10] suggests we foster a collaborative relationship between Human and AI rather than replacing composers.

3.4. Future direction

Despite the large progress made on AI's capabilities as composers, it still has great room to improve. Many current models are not adaptive: limited to specific style or genre [12]. Biaxl, an end-to-end generative model, uses MIDI data as input and tracks as output [12]. Still, current models are resource heavy: improvements can be made by optimizing training and offering diverse samples. Integrating existing music theories through data will enhance benefit the generation. End-to-end music composition can be further evaluated. Models implementing DDSP should be further investigated as it has shown the ability of extrapolating information from existing data as well as using new models that follow timbre transcript [5]. Advancing DDSP could lead to generation of higher fidelity audio [5]. The actual supportive capability of AI is nonetheless important when discussing artificial intelligence's role in music composition. More research should be done to examine AI's supportive capabilities: its ability to cocreate with composers [10].

4. Conclusion

Current generative models are rigid in their output and the music composed often lacks certain traits such as timbre and clarity especially when working with long time. Differentiable Digital Signal Processing automizes parameter tuning and addresses many limitations faced by traditional neural models: waveform alignment, spectral leaks, lack of bias.

An end-to-end approach for generating music integrates many traditional steps of music composition into one internal process, giving the AI more controller over each element. Vector representations of elements enables a clearer relationships between variables.

By function, AI aims to support the composer as a co-creator rather than a tool; though current understanding conceptualizes it as a generational resource. It adds onto the existing composing process with a new step of interacting and combining AI generated content with personal composing, allowing novice composers to be more familiar and confident in practice.

The progress of artificial intelligence should not overshadow the challenges it still faces. AIs struggle with "humanness" also appears in music composition. Questioning over its creativity and

originality arises. Copyright laws debate over whether AI generated content deserves ownership, which is ultimately dependent on whether it demonstrates a user's creativity and artistic intent.

This paper presents a review of the current position of artificial intelligence in music composition. This evaluation is based upon the rigorous academic papers published in recent time. Further exploration of this topic can be done by doing my own empirical research. By personally testing different models and their variations, I can collect data on the effectiveness of AI in controlling specific variables like timbre and pace in music composition. Overall, this review reinforces AI as a powerful developing tool that may serve as a supporting companion in music composition.

References

[1] Gilbert É, Conklin D. A probabilistic context-free grammar for melodic reduction. InProceedings of the International Workshop on Artificial Intelligence and Music, 20th International Joint Conference on Artificial Intelligence 2007 Jan 6 (pp. 83-94).

[2] Bel B, Kippen J. Modelling music with grammars: formal language representation in the Bol Processor. Computer representations and models in music. 1992: 207-38.

[3] Liang Q, Zeng Y, Tang M. Mode-conditioned music learning and composition: a spiking neural network inspired by neuroscience and psychology. arXiv preprint arXiv: 2411.14773. 2024 Nov 22.

[4] Robert-Constantin I, Trăușan-Matu S. A quantitative aesthetic analysis of artificial intelligence generated music. Proceedings or RoCHI. 2023: 63-8.

[5] Engel J, Hantrakul L, Gu C, Roberts A. DDSP: Differentiable digital signal processing. arXiv preprint arXiv: 2001.04643. 2020 Jan 14.

[6] MIDI-DDSP: Detailed Control of Musical Performance via Hierarchical Modeling

[7] Wu Y, Manilow E, Deng Y, Swavely R, Kastner K, Cooijmans T, Courville A, Huang CZ, Engel J. MIDI-DDSP: Detailed control of musical performance via hierarchical modeling. arXiv preprint arXiv: 2112.09312. 2021 Dec 17.

[8] Caspe F, McPherson A, Sandler M. DDX7: Differentiable FM synthesis of musical instrument sounds. arXiv preprint arXiv: 2208.06169. 2022 Aug 12.

[9] Rohrmeier M. On creativity, music's AI completeness, and four challenges for artificial musical creativity. Transactions of the International Society for Music Information Retrieval. 2022 Mar 9; 5(1).

[10] Civit M, Civit-Masot J, Cuadrado F, Escalona MJ. A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. Expert Systems with Applications. 2022 Dec 15; 209: 118190.

[11] Gianet ET, Di Caro L, Rapp A. Music Composition as a Lens for Understanding Human-AI Collaboration. InCEUR WORKSHOP PROCEEDINGS 2024 (Vol. 3701, pp. 1-7). CEUR-WS.

[12] Hayes B, Shier J, Fazekas G, McPherson A, Saitis C. A review of differentiable digital signal processing for music and speech synthesis. Frontiers in Signal Processing. 2024 Jan 11; 3: 1284100.

[13] Mao HH, Shin T, Cottrell G. DeepJ: Style-specific music generation. In2018 IEEE 12th international conference on semantic computing (ICSC) 2018 Jan 31 (pp. 377-382). IEEE.

[14] Madhumani GR, Yu Y, Harscoët F, Canales S, Tang S. Automatic neural lyrics and melody composition. arXiv preprint arXiv: 2011.06380. 2020 Nov 12.

[15] Ferreira P, Limongi R, Fávero LP. Generating music with data: application of deep learning models for symbolic music composition. Applied Sciences. 2023 Apr 3; 13(7): 4543.

[16] Hernandez-Olivan C, Hernandez-Olivan J, Beltran JR. A survey on artificial intelligence for music generation: Agents, domains and perspectives. arXiv preprint arXiv: 2210.13944. 2022 Oct 25.

[17] Mycka J, Mańdziuk J. Artificial intelligence in music: recent trends and challenges. Neural Computing and Applications. 2024 Nov 16: 1-39.

[18] Bulayenko O, Quintais JP, Gervais DJ, Poort J. AI music outputs: Challenges to the copyright legal framework. Available at SSRN 4072806. 2022 Feb 28.

[19] Fu Y, Newman M, Going L, Feng Q, Lee JH. Exploring the Collaborative Co-Creation Process with AI: A Case Study in Novice Music Production. arXiv preprint arXiv: 2501.15276. 2025 Jan 25.