Tacit and Explicit: the Structural Dilemma of Generative Artificial Intelligence Music Style Cognition Driven by Prompts

Jiaqian Chen[1]

[1]*Communication University of China, Beijing, China*
[1]*Email: 2023101301j3137@cuc.edu.cn*

ABSTRACT
Prompts-driven generative artificial intelligence music uses natural language as the explicit interactive medium, attempts to translate the highly context-dependent musical style attributes that rely on tacit cognition into computable textual instructions. This tacit-to-explicit translation assumption itself constitutes the root cause of the structural dilemma. This thesis bases on the tacit cognition theory proposed by Polanyi, and investigates the mechanism and challenges of the shift from tacit to explicit of artificial intelligence music styles cognition in the prompts interface. It points out that the structural position differences of tacitness in human cognition and generative artificial intelligence systems are the key factors that exacerbate the predicament. The prompts interface compresses the structure of tacit and explicit interweaving in human music style cognition into the unidirectional reliance on the explicit language. It respectively causes the reductive dilemma in language translation, the homogenization dilemma of style standards, and the disembodiment dilemma of the creative subject in three aspects of language input, model generation and creative subject. The generative artificial intelligence music's development should go beyond the simplistic thinking of technical optimization, attach importance to the epistemological resolution of structural dislocation, reserve an irreplaceable space for human tacit participation, make the output of artificial intelligence a trigger mechanism that stimulates rather than replaces human tacit participation, and promote the human-computer music creation to move from superficial style replication to deep aesthetic collaboration.

Keywords: *Generative artificial intelligence music, Prompts, Tacit cognition theory, Style cognition, Structural dilemma.*

1. INTRODUCTION

The generative artificial intelligence (GenAI) music driven by prompts is the key paradigm in the evolution of artificial intelligence music creation from rule-based algorithmic composition and statistical learning-based generative models to deep learning-based cross-modal generation systems [1]. Nowadays, the text-to-audio models represented by Stable Audio Open, Suno and Udio presuppose natural language as an effective proxy of structured latent representations, and reduce the music styles as explicit words that can be spelled out. But this assumption is on the contrary of the core viewpoint in the field of music for a long time: music cannot be completely described by language, it is also not equal to language, but the thing that goes beyond the limit of the pure structure [2][3]. The GenAI music style cognition research driven by the prompts, as the important constituent part of the interdisciplinary field of music cognition, artificial intelligence and art theory, has the epistemological logic and theoretical significance distinguished from the research on traditional music cognition and the research on pure technical music generation. Its profundity of study will directly influence the completeness of the theoretical framework for human-computer collaborative music creation [4]. This special value determines the crucial role that style cognition plays in the research of GenAI

music. It can be said that it not only is the intuitive examination of the translation effect of human music aesthetic experience into machine systems, but also reflects the explanatory efficacy of the art cognition theory in the digital creative field.

In recent years, although the GenAI music system driven by prompts has expanded the granularity of style control from macro genre labels to high-level attributes such as instrument arrangement, emotional atmosphere, and musical form segments [5]. However, for the fine-grained, temporal tacit dimensions that constitute the core of the music style, the prompts interface still faces an insurmountable "explanation gap" [6]. This gap is not a functional gap that can be filled by technological iterations, but the structural tension between tacit knowledge and explicit representation in the cognitive understanding of musical styles that constitutes a fundamental issue that requires inquiry from the perspective of epistemology. This thesis takes Polanyi's tacit cognition theory as the overarching framework to investigate the transformation mechanism of style cognition in GenAI music driven by prompts, as well as its epistemological consequences. It summarizes the unfolding of this transformation mechanism as three levels of dilemmas: language input, model generation, and the creative subject. It also reveals that the root cause of these dilemmas lies in the structural mismatch between tacit and explicit elements. This thesis expects to provide theoretical reference for the research of GenAI music's style cognition, also to provide theoretical enlightenment for conducting the research of AI music aesthetic cognition in the fields of art technology and digital humanities.

2. TACIT AND EXPLICIT: THE DUAL ASPECTS OF MUSIC STYLE COGNITION

2.1 The Tacit Cognition Theory and Its Music Applicability.

The tacit cognition means the activity or process of "knowing", that is the cognitive method for us to integrate various details to understand things, but unable to state the specific basis. In Michael Polanyi's words, it is the underlying mechanism of the phenomenon of "we can know more than we can tell" [7]. Polanyi's distinction of tacit knowledge and explicit knowledge constitutes the fundamental category for understanding the structure of human knowledge. The so-called tacit knowledge refers to those forms of knowledge that cannot be fully expressed in words but actually govern people's cognition and actions; while explicit knowledge refers to the knowledge that can be explicitly expressed and conveyed in the form of language, symbols, formulas, etc. At the same time, "the process of formalizing all knowledge to the exclusion of any tacit knowing is self-defeating" [8], the explicit knowledge must rely on tacit cognition to be applied and transmitted.

Polanyi also uses the distinction of "subsidiary awareness" and "focal awareness" to show the structure of tacit cognition [9]. The relationship between these two types of awareness should be respectively understood from the perspectives of functional roles and integrated operations. In terms of functional roles, the two are mutually exclusive: the same set of details cannot be examined in a focused manner and relied upon in an auxiliary way simultaneously. Take the piano playing as an example, once the player turns the focus of attention to the details of the finger movements, the finger no longer serves as subsidiary cues for integration, and as a result, the overall performance of the music will collapse. But in normal integration of operations, the two are functionally complementary. When the player focuses on the overall music, the tactile sensation and movement of the fingers precisely serve as dependent subsidiary cues in the integration process. The subsidiary awareness provides detailed support for the focal awareness, while the focal awareness offers an integration direction for the subsidiary details.

Since the tacit cognition theory of Polanyi was proposed, it has been widely applied in many fields like music education, music performance and music cognition etc. For example, Bowman introduced Polanyi's tacit cognition theory into the field of music education, clearly stating that music learning is essentially the construction of tacit knowledge. He also demonstrated that the tacit cognition theory can explain how formal music learning affects musical aesthetic experience [10].Mitchell et al. pointed out from the perspective of multi-sensory music performances that the tacit knowledge plays an important role in cross-modal integration of music performances [11]. Pearce's statistical learning research shows that music style cognition relies on the listener unconsciously internalizing the statistical structures inherent in their specific musical culture [12]. While, the above researches mainly focus on the tacit cognition mechanism in human music learning and performance. The

research on the structural dislocation between tacit and explicit structures in GenAI music style driven by prompts is still lacking. Considering that the style cognition of GenAI music is a complex systematical question which involves cognitive mechanisms, technical representations and knowledge types, the analysis of it should use the theoretical tools that can take into account both the cognitive process and the knowledge structure. This research refers to Polanyi's tacit cognition theory to analyze GenAI music style cognition. It not only considers the technical reality of the prompts interface translating tacit knowledge into explicit language labels, but also regards the translation process itself as the object of epistemological analysis. This not only can better provide theoretical reference for style cognition research of GenAI music, but also can reveal the structural position difference of tacit cognition and explicit representation in cognition systems of human and AI.

2.2 The Music Style: the Interwoven Area Between Tacit Knowledge and Explicit Knowledge.

Human's cognition for music style involves the dynamic interaction between tacit participation and explicit representation. On the one hand, music style has explicitizability and describability. It can be described as the quantifiable structural elements, such as melodic features, harmonic vocabulary, rhythmic patterns, orchestration methods, tonal scales, genre conventions. These elements can be named and classified by musical theoretical language, and can serve as explicit references for style identification and style teaching. On the other hand, the deep meaning of music style is far more than the mere sum of these explicit characteristics. Its vitality engages in embodied musical practice. The player's "feeling" of a certain melody, the jazz musician's "swing feeling" in improvisation and the untranslatable "charm" of ethnic music all involve Polanyi's concepts of the tacit dimensions that indwelling in the body and the context.

Polanyi pointed out that this action of "indwelling" has specific logical structure. The tacit cognition always includes two terms: we attend from some things to attend to other things. The former one is called the "proximal term", and the latter one is called the "distal term" [13]. The proximal term is the detail, tool or body movement of which we are subsidiarily aware. The distal term is the whole meaning or objects that we focally pay attention to. In musical practice, the player's body movement, instrument touch and auditory details constitute the proximal term, and the whole cognition and presentation of music style constitute the distal term. It is precisely by indwelling themselves in these embodied details that the player is able to grasp the overall meaning of the music style. Rohrmeier and Rebuschat pointed out that music knowledge (including music style) is not explicit knowledge to a large extent, but is mainly a kind of implicit and unconscious knowledge. It is acquired through exposure to and interaction with many music samples, and is usually cannot be completely explained [14]. The music identify tasks like style identify, forecast etc. rely on the syntactic knowledge which dominates the music style, but we usually cannot provide reasons for one's own judgment. Therefore, the music style should not only consist nameable explicit layer, but also deeply roots in not fully describable tacit layer. Its complete cognition and creation must simultaneously engage both of the two levels.

It is just this interwoven nature of the knowledge structure that enables the music style to simultaneously possess both an explicit aspect that can be captured by language labels, and a tacit aspect that cannot be fully conveyed by language. However, when GenAI music systems use prompts as intermediary to "study" and "generate" styles, what it relies on is precisely the explicit aspect of music style. However, what this reliance means in structure is something that requires an analysis of the GenAI music style generation mechanism driven by prompts.

2.3 The GenAI Music Style Generation Mechanism Driven by Prompts

The core technology mechanism of GenAI music driven by prompts is that it uses text embeddings as the conditional signal, and guides the diffusion model to gradually generate music in a low-dimensional latent space. In this mechanism, user's music style intension is turned into explicit language symbols, and then encoded as feature vectors, ultimately "executed" as acoustic forms. Examining this mechanism through Polanyi's distinction between "subsidiary awareness - focal awareness", its cognitive meaning becomes apparent. The two kinds of awareness are mutually exclusive in terms of their functional roles. The same set of details cannot simultaneously be focally scanned and be subsidiarily relied upon. However, the prompts interface asks the user to "translate"

the style intention into a language tag, which forcibly brings the embodied details that were originally at the level of subsidiary awareness to the focal level.

Messina et al. provide empirical support for this judgment. They point out that the tacit culture knowledge, as an internalized cultural repository, is circulated and manifested through the construction of textual cues and the evaluation of LLM generated audio [15].Different from the knob labels on the analog synthesizer used to describe measurable parameters, the language of AI texts-to-sound applications usually stops at more everyday cultural level. Park et al. further reveal that any notation system will give priority to the content that can be expressed easily, while flattening the content that cannot be expressed; although the prompts can conveniently assign high-level attributes such as emotion, instrument arrangement and style reference, it is difficult to clearly express the microscopic level where music resides and the decisions based on time [16]. In the process that is driven by prompts, many musical judgments that originally existed in the microscopic temporal details have been transferred to the explanatory layer of system management, and as a result, the control has become asymmetric. This is because the necessary tacit knowledge and explicit interwoven in the cognition of music styles are replaced under this condition by a one-way dependence on explicit language.

It should be noted that language itself has tacit dimensions. Polanyi pointed out that “When we use words in speech or writing we are aware of them only in a subsidiary manner.”[17] This means that the usage of language itself relies on tacit integration, while the prompts simplify language into explicit labels, which ignore the tacit integration that the usage of language itself relies on. When the tacit knowledge contained in music style is required to go through the explicit language channel in order to enter the generation process, the most vital dimension of it is at risk of being structurally obscured. This risk is not the functional weakness in the specific period of technology development, but the specific manifestation of the incompatibility in type between the preset “one-dimensional explicitization” cognitive path in the prompts interface and the tacit structure of musical style cognition.

3. THE STRUCTURAL DISLOCATION: THE THREEFOLD DILEMMA IN THE OF GENAI MUSIC STYLE COGNITION DRIVEN BY PROMPTS

The origin of GenAI music style cognition dilemma driven by prompts lies in the structural displacement between tacit and explicit. The prompts interface compresses the organically interwoven structure of the tacit knowledge and explicit knowledge in music style cognition into a one-way reliance on explicit language. This thus triggered an incommensurable epistemological dislocation at the three levels of language input, model generation, and the creative subject. The following three dilemmas are the unfolded forms of this structural displacement in different level.

3.1 The Reductive Dilemma at the Language Input Level

The reductive dilemma of language tranlation is the manifestation of this structural displacement in language input level. When the continuous, embodied and context-dependent features of a musical style are compressed into discrete, labelable language tags by the prompts, the most subtle tacit connotations within the style will inevitably be flattened, simplified or omitted during the process of expression. Many core elements in music style, such as the subtle luster of the tone, the breath-like quality of the musical phrases, and the swinging tension in the improvisation, are difficult to accurately convey through language. The essence of language is discrete and categorized; while music style’s tacit level is continuous and indivisible.

This reductibility in music creation practice shows as a significant gap in expression. The empirical study conducted by Choi et al. revealed that although the user groups use the GenAI music system based on prompts for different purposes: experts use it to verify musical concepts, beginners use it to generate reference samples, and non-professionals use it to transform abstract ideas into musical works, all users are confronted with the limitations of the prompts in handling temporal aspects and subtle musical differences [18]. In other words, there is an insurmountable gap between the style information that the prompts can convey and the actual cognitive structure of musical styles.

It is worth noting that the depth of this dilemma is not only the expression loss on the human side. The study of Singh et al. based on sparse autoencoders extracts interpretable features from the residual flow of the autoregressive music generator MusicGen, revealing two coexisting patterns: the first one is familiar music concept, which means features that can be clearly corresponded to existing music theories or everyday language; the second one is a coherent but unencoded pattern that lacks a clear counterpart within existing theories or languages [19]. The latter one shows that there are also structural characteristics that resist language description in GenAI music model, which can conceptualize this type of pattern as "computational unencoded residuals". It has similarity with human tacit knowledge in the perspective of function, which means the two both resist complete externalization, but it is totally different in terms of existence mode: human tacit knowledge emerges in embodied participation, while the computational unencoded residuals are accumulated during the statistical training process. It needs to be emphasized that this concept is a kind of functional analogy, but not ontological identity. Polanyi's concept of "tacit" is closely related to physicality, embodiment and active principle, while the statistical residuals of AI are not associated with any physical experience or active principles. This discovery transforms the inaccuracy in the imprecision of linguistic translation into a structural dislocation between the two types of knowledge. Even though the prompts can more accurately specify explicit style features, there are still tacit dimensions that cannot be fully utilized by language inside the model.

3.2 The Homogenization Dilemma at the Model Generation Level

This structural displacement extends to the model generation level, manifesting as the homogenization dilemma of style standard. The explicitization logic of prompts not only acts upon individual creator's expression, but also deeply reshapes the collective perception of what constitutes a generative style. The systematic training data from text to audio is mainly collected and curated by a few tech giants. The style map leaned by the model is actually a standardized projection of the dominant sound culture represented by these data.

Slendebroke and Metaxa provide empirical evidence from the computational level for Suno and Lyria 3's systematic audit. This research uses the black-box deployment assessment method, without accessing the model training data and architecture, the performance of two commercial deployed text-to-music systems in four genres was audited. A comparative analysis was conducted based on 72 music information retrieval features. The research finds that the two kinds of systems show two homogenizing trends that differ in structure: Lyria 3 reduces the acoustic diversity within the genre, while Suno reduces the acoustic distinctions between different genres. What is more notable is that the research finds that neither of the two systems strictly followed the user's prompts. The observed patterns reflect the prior knowledge acquired by the model rather than the constraints imposed by the prompts [20]. This means the root cause of homogenization lies not in how users utilize the prompts words, but in the fact that the style representations encoded by the model have systematically favored the dominant and explicit style paradigm.

This deep mechanism of homogenization trend lies in the fact that explicitization inherently carries a tendency towards standardization. Polanyi has previously indicated that the process of making tacit knowledge explicit often involves the extraction of individualized and context-specific content, and instead replaces it with universalizable and replicable categories. In the context of GenAI music, when the style is made explicit by the prompts, it is also included in a style possibility space that is jointly defined by data distribution, embedding space, and platform mechanisms at the same time. Mehta et al. analyzed an audio dataset for AI-generated music with over one million hours of data, and found that approximately 86% of the total dataset duration comes from music from the Global North, while the Global South genres account for only 14.6%. [21]. The research of Solak et al. further shows that the most advanced music generation models have significant biases across different regions and genres. Music produced in low-resource areas and by niche genres is more likely to have outliers in its distribution [22]. The skewness of data distribution is not technology-neutral; it directly shapes the cognitive boundaries of the model regarding the recognition of musical styles. If things continue this way, the music styles generated by the creators under the influence of the prompts will increasingly converge towards the combination of the explicit style paradigms encoded by the model. The unique style sounds

which truly emerge from embodied tacit knowledge face the risk of being diluted and disappearing.

3.3 The Disembodiment Dilemma at the Creative Subject Level

The dilemma of disembodiment at the level of the creative subject reveals the result of this structural displacement on the subject level. The GenAI music driven by prompts places the creator in front of a language-centered interactive interface. His body just acts as the typing hand, but not as the embodied media for style expressing. The traditional music composition and performance are highly embodied activities. The composer's style purpose realizes through the indwelling of musical instruments, voices and body movements. The creator keeps direct, kinesthetic, sensorimotor connection with sound. However, in GenAI music driven by prompts, this embodied loop is suspended by the language interface.

The disembodiment consequences of GenAI music creation driven by prompts is multiple. First, the absence of embodied feedback erodes the creator's sense of activity, and as a result, the style intention becomes impoverished. A research points out from the perspective of the new Deweyan pragmatism that the art forms like music are totally embodied human experience. It involves human activity, the phenomenology of aesthetics, and the mutual recognition between artists and audiences. The research also warns the easy, agency-eroding pleasures that GenAI music may bring to creators [23]. Second, the creator's author identity and sense of belonging for the generation result are weakened, and then trigger deep anxiety about the author identity. Sturm conducted a self-reflection study based on his own use of Suno and Udio for creating prompts music through LLM intermediary. It reveals that when the creators are negotiating author identity and confronting the sense of fraud, they at the same time are constantly encountering the ambiguity and reconfiguration of the creative subjectivity during the generation process. Third, from a more profound philosophical perspective, when AI can constantly generate music without the creator's personal involvement and concrete life experience, this has already touched upon the issue of creative subjectivity itself. Zhao proposed the concept of "authorial positioning", which means the creators maintain the presence of the author and the creative subjectivity through reflective and strategic practices under the conditions of the algorithm [25]. This research discovers that the creators generally experience a kind of ambiguous sense of authorship that is "neither completely present nor completely absent". This identity could neither be naturally conferred by the act of creation, nor be obtained through complete control of the generation process, but must rely on repeated identification, evaluation and adjustment to constantly confirm. It indicates that when music can be constantly generated without the creator's direct involvement in concrete generation step (even lack of the traditional sense of "personal creation"), the identification relationship between the work and the creative subject is no longer self-evident, but become a process that requires repeated negotiation and re-establishment. This reminds that the disembodiment may not only give rise to technical or ethical issues, but is more likely to involve a loosening and reconfiguration of the relationship between music as a symbol and its creative subject.

4. CONCLUSION

The style cognition dilemma of GenAI music driven by prompts essentially are not functional defects at specific stages of technological development, but a kind of structural epistemological displacement revealed by the perspective of Polanyi's tacit cognition theory. The origin of this displacement lies in the fact that the prompts interface forcibly compresses the organic interweaving structure of tacit and explicit knowledge in human music style cognition into a one-way reliance on explicit language. As a result, it triggers an incommensurable epistemological displacement at the three levels of language input, model generation, and the creative subject. The irreducibility of human tacit knowledge is not a kind of limitation that needs to be "overcome", but the principled difference existing between the conditions themselves of human creation and the structures of the calculating system.

The revelation of this structural dilemma has direct theoretical implications for the movement of human-computer music creation from "surface style replication" to "deep aesthetic collaboration". Grégoire et al., based on their ethnographic observations of music creators' use of GenAI tools in recording studio environments, pointed out that AI functions primarily as an inspirational tool and collaboration catalyst, rather than replacing creators [26]. However, this research does not inquire about what conditions AI needs to fulfill in order to function as a catalyst for collaboration from the perspective of epistemology. The structural

dilemma analysis exactly responds to this ongoing issue. The possibility of AI collaborative catalysis lies precisely in the fact that human creators maintain a dimension of tacit participation beyond the prompts interface. If the prompts interface excludes the operating space of subsidiary awareness structurally, then the collaborative catalysis may degenerate into superficial style imitation. The research of Hopkins et al. further reveals the technological dimension of this problem. The collaboration among improvisational musicians relies on the dynamic exchange of subtle differences in human musical communication. Many musicians can intuitively perceive this information, but it is still a persistent challenge to convert this knowledge into an embodied computer-driven music system, whether it is a robot or a virtual musician [27]. This judgment and the structural dilemma analysis in this thesis form a direct dialogue. AI system's management of tacit knowledge still stops at the statistical pattern extraction level, it cannot match the tacit integration that emerges in real-time through embodied participation in human improvisational collaboration.

The dilemma also contains the chance of transformation. The three dilemmas revealed in this thesis can provide enlightenment for the design practice of GenAI music system from three aspects. First, from "completely transform" to "admit the limit". The reductive dilemma of language translation hints that the design of prompts interface should not aim at "eliminating the tacit dimension", but on the premise of acknowledging the structural displacement, reserve space for the style connotations that cannot be captured by language. Second, from "single label" to "multimodal perception". The homogenization dilemma of standard style implies that expands the perceptual channels for stylistic semantics, that is, incorporates multimodal cues such as images, actions, and tactile sensations into the generation conditions, can help reduce the reliance on the single language category, and thereby to some extent compensates for the reductive loss in language translation. Third, from "language instructions" to "embodied participation". The disembodiment dilemma of creation subject reveals that design of the generation interface should reintroduce the physical movement, gestures and kinesthetic feedback, so the that tacit knowledge can gain an expression channel beyond language, enabling creators to maintain the embodied participation dimension beyond the language channel.

REFERENCES

[1] Zhao, Y., Yang, M., Lin, Y., Zhang, X., Shi, F., Wang, Z., Ding, J., & Ning, H. (2025). AI-Enabled Text-to-Music Generation: A Comprehensive Review of Methods, Frameworks, and Future Directions. *Electronics*, *14*(6), 1197. https://doi.org/10.3390/electronics14061197

[2] Morreale, F., Martinez-Ramirez, M.A., Masu, R., Liao, W. & Mitsufuji, Y. (2025). Reductive, Exclusionary, Normalising: The Limits of Generative AI Music. *Transactions of the International Society for Music Information Retrieval*, 8(1), 300-312.https://doi.org/10.5334/tismir.256

[3] Lidov, D. (2005). *Is language a music?: Writings on musical form and signification*. Indiana University Press.

[4] Prithvi Aravind. Toward Style-Aware Embeddings in AI-Driven Music Production: A Survey and Future Directions. *TechRxiv*. 06 November 2025.https://doi.org/10.36227/techrxiv.176240574.44870483/v1

[5] Melechovsky, J., Guo, Z., Ghosal, D., Majumder, N., Herremans, D., & Poria, S. (2024, June). Mustango: Toward controllable text-to-music generation. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)* (pp. 8293-8316).https://doi.org/10.18653/v1/2024.naacl-long.459

[6] Zang, Y., & Zhang, Y. (2024). The Interpretation Gap in Text-to-Music Generation Models. *ArXiv, abs/2407.10328*.https://doi.org/10.48550/arXiv.2407.10328

[7] Polanyi, M., & Sen, A. (2009). *The tacit dimension*. University of Chicago press.p4.

[8] Polanyi, M., & Sen, A. (2009). *The tacit dimension*. University of Chicago press.p20.

[9] Polanyi, M. (1958). *Personal knowledge: towards a post-critical philosophy*. University of Chicago press.p55-57.

[10] Bowman, W. D. (1982). Polanyi and Instructional Method in Music. *Journal of*

Aesthetic Education, 16(2), 75–86. https://doi.org/10.2307/3332279

[11] Mitchell, H. F., Blom, D., & Long, P. (2026). 'We hear with our eyes!' Unlocking tacit knowledge about multisensory music performing. *International Journal of Music Education*, 44(1), 174-187. https://doi.org/10.1177/02557614241264900

[12] Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. *Annals of the new York Academy of Sciences*, 1423(1), 378-395.https://doi.org/10.1111/nyas.13654

[13] Polanyi, M., & Sen, A. (2009). *The tacit dimension*. University of Chicago press.p10-16.

[14] Rohrmeier, M., & Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in cognitive science*, 4(4), 525-553.https://doi.org/10.1111/j.1756-8765.2012.01223.x

[15] Messina, M., Aliel, L., Keller, D., & Gómez Mejía, C. M. (2026). Actionable tacit knowledge in second-wave ubiquitous music frameworks. *AI & SOCIETY*, 41(3), 2345-2360.https://doi.org/10.1007/s00146-025-02471-y

[16] Park, E. J. (2026). The missing middle in AI native music production: Compression, editorial musicianship, and governance in the generative stack. *AI Magazine*, 47(3), e70072.https://doi.org/10.1002/aaai.70072

[17] Polanyi, M. (1958). *Personal knowledge: towards a post-critical philosophy*. University of Chicago press.p57.

[18] Choi, Y., Moon, J., Yoo, J., & Hong, J. H. (2025, April). Understanding the potentials and limitations of prompts-based music generative AI. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-15).https://doi.org/10.1145/3706598.3713762

[19] Singh, N., Cherep, M., & Maes, P. (2026, April). Discovering and steering interpretable concepts in large generative music models. In *International Conference on Learning Representations* (Vol. 2026, pp. 84922-84941).https://doi.org/10.48550/arXiv.2505.18186

[20] Slendebroek, Z., & Metaxa, D. (2026). The Algorithmic Flattening of Sound: Computational Evidence and Justice Implications of AI Music Homogenization. *Proceedings of the AAAI ACM Conference on AI, Ethics, and Society*, 9(4), 2353–2364. https://doi.org/10.1609/aies.v9i4.43312

[21] Mehta, A., Chauhan, S., & Choudhury, M. (2026). Missing Melodies: AI Music Generation and Its "Nearly" Complete Omission of the Global South. *Proceedings of the AAAI ACM Conference on AI, Ethics, and Society*, 9(3), 1810–1819. https://doi.org/10.1609/aies.v9i3.43269

[22] Solak, Ahmet, et al. "Bias beyond Borders: Global Inequalities in AI-Generated Music." *arXiv preprint arXiv:2510.01963 (2025)*.https://doi.org/10.48550/arXiv.2510.01963

[23] Casey Haskins, Is AI a Threat to Human Music? (a neo-Deweyan Perspective), *The Journal of Aesthetics and Art Criticism*, Volume 83, Issue 3, Summer 2025, Pages 238–253, https://doi.org/10.1093/jaac/kpaf024

[24] Sturm, B. L. (2025). " I made this (sort of)": Negotiating authorship, confronting fraudulence, and exploring new musical spaces with prompts-based AI music generation. *arXiv preprint arXiv:2507.23365*.https://doi.org/10.48550/arXiv.2507.23365

[25] Zhao, S. (2026). Authorial Positioning and Creative Agency in AI-Mediated Music. *Global Media and China*, 11(3), 352-367.https://doi.org/10.1177/20594364261423866

[26] Grégoire, A., Lacasse, S., & De Blois, J. V. (2026, April). Generative Artificial Intelligence in Music: Towards an Aesthetics of Co-creation. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)* (pp. 97-112). Cham: Springer Nature Switzerland.https://doi.org/10.1007/978-3-032-24350-8_7

[27] Hopkins, T., Sun, R., Chuan Weng, S. C., Ma, S. Y., Crum, J., Hirshfield, L., & Do, E. Y. L. (2025). Brain jam: Neural signal-informed adaptive system for drumming collaboration with an ai-driven virtual musician. *Frontiers*

in Computer Science, 7, 1570249. https://doi.org/10.3389/fcomp.2025.1570249