Chinese Characters as Digital Scores: A Prototype for Explainable, Semantics-Driven Composition

Zhang Yuan
zhang_yuan@ccom.edu.cn

Zhang Xinran
zxinran@ccom.edu.cn

ABSTRACT

This paper presents an explainable semantic-driven composition framework that generates controllable sound code from cultural symbols. Unlike black-box AI music generation systems that focus primarily on stylistic imitation, our approach leverages semantic information—including linguistic meaning and cultural context—from a large language model (LLM) to guide music creation. Building upon and extending the Digital Score [1] paradigm, our framework treats Chinese characters not merely as symbolic inputs but as semantic carriers forming a unique modality of music notation: their semantic and structural features are algorithmically mapped into performable musical instructions. This enables interpretable relationships between symbolic meaning and sonic expression, positioning the character itself as a carrier of musical intention. A prototype system is implemented that maps semantic embeddings of Chinese characters to sonic parameters and symbolic score structures, and implements a prototype capable of producing executable sound code. The prototype allows users to trace, visualize, and adjust semantic-to-sonic mappings, improving creative control and cultural representation. An artistic case study demonstrates how Chinese characters can be transformed into interactive, interpretable digital scores, highlighting the potential of semantic-driven composition to foster transparency, cultural integration, and new modes of creative agency in algorithmic music practices.

1. INTRODUCTION

Artificial intelligence (AI) has significantly transformed music composition, offering novel ways to generate sound materials, automate workflows, and explore creative boundaries beyond human limitations. Yet many AI-driven systems operate as "black boxes," focusing on stylistic imitation or statistical pattern generation with limited interpretability. This opacity raises challenges for composers and researchers, particularly when artistic control and cultural representation are at stake.

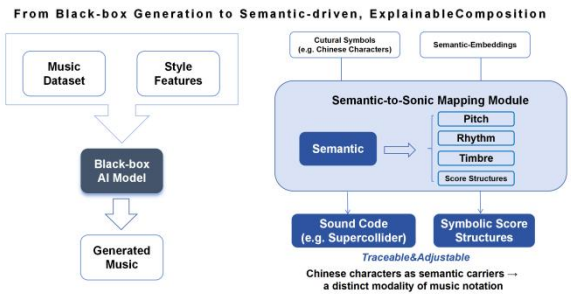

Figure 1. Comparison between black-box AI music generation and our semantic-driven, explainable composition framework, highlighting Chinese characters as semantic carriers forming a unique modality of music notation.

Digital Score, as conceptualized in recent research, has reimagined notation as an interactive, computationally mediated space for creative performance. This work builds upon and extends that concept by introducing a semantic-driven approach [1, 2] in which Chinese characters themselves function as semantic carriers forming a unique modality of music notation. Their semantic and graphic structures are computationally interpreted as performable musical instructions, enabling the character to serve as both cultural sign and music notation. Instead of relying solely on numerical features or stylistic references, our framework integrates semantic embeddings and algorithmic mappings by leveraging the powerful semantic modeling functionality from a large language model (LLM), allowing users to trace and control how meaning informs sound generation.

We present a prototype implementation using Chinese characters as a case study. Semantic features derived from the LLM for the linguistic and cultural attributes are mapped to sonic parameters and symbolic score structures, producing sound code executable in SuperCollider. In doing so, the prototype reframes the Chinese character as a performable and interpretable digital score, enriching the conceptual space of notation by merging linguistic semantics with musical structure. This design offers explainable semantic-to-sonic mappings and supports creative agency.

The contributions of this paper are threefold: (1) it proposes a framework for explainable, semantic-driven composition that extends Digital Score concepts; (2) it implements a prototype that links cultural symbols—here, Chinese characters — to sound code as performable notation; and (3) it presents an artistic case study illustrating how semantic meaning can be transformed into interpretable musical structures using LLM. These contributions aim to bridge semantic representation, cultural expression, and explainability within algorithmic music composition.

2. AI COMPOSITION: SEMANTICS, EXPLAINABILITY, AND CULTURAL SYMBOLS

Artificial intelligence has reshaped compositional practice, providing automated generation systems capable of producing musical material in a wide range of styles. Early approaches focused primarily on statistical models and style imitation, including Markov models, genetic algorithms, and, more recently, deep generative models such as recurrent neural networks and Transformers. While these systems successfully capture stylistic features, they often function as "black boxes," offering limited insight into how input data shapes the resulting output. This limitation has stimulated interest in semantic-driven [2, 3] approaches, where the goal is not simply to reproduce stylistic surfaces but to integrate meaningful relationships between symbolic input and sonic output. Existing research in semantic composition explores emotional mapping, text-to-music transformation, and cross-modal correspondences, but has rarely considered cultural symbols as structured semantic carriers.

The rise of Explainable AI (XAI) has brought new possibilities for music technology, emphasizing transparency and user control. In composition systems, explainability allows composers to trace algorithmic decisions, adjust creative parameters, and integrate machine-generated content within coherent artistic visions. Most XAI research in music, however, focuses on model interpretation or user interfaces, rather than on the semantic transparency of the compositional material itself. Our approach extends this discussion by proposing that semantic meaning — encoded as structured embeddings from cultural symbols—can become a core element of an explainable compositional process.

The use of cultural symbols in music is not new: calligraphic scores, poetic transformations, and visual music traditions have long sought to merge language, image, and sound [4, 5]. In digital media art, these ideas have evolved into cross-modal mapping techniques, including graphical scores, live visuals, and multimodal generative systems. However, most existing work treats symbols either as aesthetic inspiration or as fixed visual references, rather than as computationally interpretable semantic carriers. By framing Chinese characters as structured semantic entities and integrating them into an algorithmic composition pipeline, our research situates itself at the intersection of cultural representation, semantic processing, and digital score innovation.

While there is growing work on AI composition, explainability, and cross-modal use of cultural symbols, these areas have seldom converged in a unified semantic-driven framework. Our system design, presented in the next section, responds to this gap.

3. SEMANTIC-DRIVEN DIGITAL SCORE FRAMEWORK

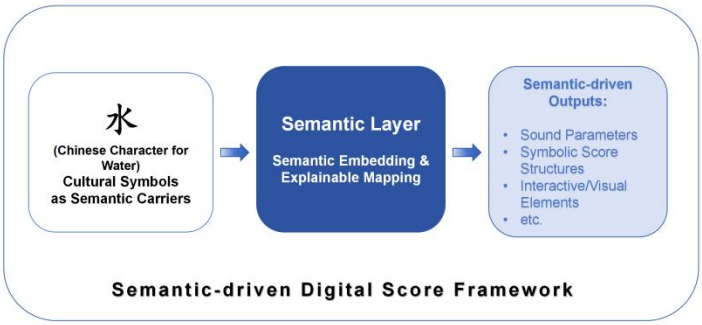

Figure 2.Semantic-driven Digital Score framework, showing how cultural symbols are processed through a semantic layer to generate diverse outputs such as sound parameters, symbolic score structures, and interactive/visual elements.

The core principle of our framework is to treat cultural symbols not as decorative triggers but as semantic carriers that can directly shape musical structure and sonic behavior. Instead of relying on opaque, style-imitating generative models, we introduce an explainable pipeline where semantic meaning flows transparently from symbol to sound. This design builds on the Digital Score concept by adding a semantic layer that translates linguistic and cultural information into computationally interpretable musical instructions. As a result, composers can trace how symbolic meaning influences pitch, timbre, and rhythm, while retaining creative control and cultural specificity.

3.1. Semantic Architecture and Processing Pipeline

The proposed framework extends the concept of the Digital Score by embedding semantic processing directly into the compositional pipeline. Instead of treating input symbols as decorative elements or fixed triggers, our system interprets cultural symbols — here, Chinese characters — as semantic carriers that influence compositional decisions. The architecture is structured around three layers: (1)semantic processing, which extracts linguistic and cultural meaning from the symbols; (2)semantic-to-sonic mapping, which translates this meaning into sound parameters and symbolic score structures; and (3)output rendering, which produces executable sound code and interpretable digital scores. This layered design ensures that symbolic meaning flows

transparently through the pipeline, resulting in a unique modality of music notation where cultural semantics directly shape musical outcomes.

At the semantic processing stage, each Chinese character is converted into its semantic embedding using an open-source LLM that has been pre-trained on Chinese text corpora rich in cultural content. This embedding captures contextual semantic nuances and associative meaning beyond the visual appearance of the character. To translate semantics into musical form, we adopt a rule-based mapping layer ensures interpretability by aligning specific semantic dimensions (e.g., "fluidity," "brightness," "motion") with musical attributes such as pitch contour, timbral quality, and rhythmic density. This allows composers to trace how symbolic feature informs sound generation while still benefiting from adaptive AI-driven mapping. The mapped parameters are then encoded as SuperCollider code, supporting direct sound synthesis, and as symbolic score structures following the Digital Score paradigm, ensuring both computational execution and performable interpretation.

3.2. Use of LLM in Semantic Processing

A key aspect of this work is the use of LLM for semantic processing and mapping. We adopt the LaBSE model [6] to convert each Chinese character into its semantic embedding. By leveraging the contextual semantic modeling power of LLM, these embeddings capture nuanced cultural and contextual representation beyond simple lexical definitions. To translate these semantic features into musical parameters, we employ a mapping function to convert the embedding into parameters that control the synthesization of music sound in SuperCollider. This LLM-driven design allows adaptive and explainable transformation of cultural semantics into sound parameters and symbolic score structures, supporting both creative flexibility and transparent control.

3.3. Cultural Extensibility

While this work is centered on Chinese characters as the primary symbolic source, their role goes beyond that of a simple demonstration. Chinese characters form a major branch within the framework, providing both a testbed for prototyping and a rich, culturally grounded symbolic system that can scale within the architecture. Their complex structure and deep semantic associations make them ideal for exploring how meaning can be algorithmically linked to musical parameters and digital score structures.

At the same time, the framework is designed for broader cultural integration. Because the semantic embedding and mapping modules are modular, they can be adapted to other symbolic systems such as Arabic calligraphy, indigenous glyphs, or abstract graphic signs. Each of these systems carries distinct visual and semantic characteristics, requiring only domain-specific embedding models and minimal training data to capture their meaning. By using multilingual or culturally specialized pre-trained LLMs, the framework can generalize semantic processing across diverse writing systems and symbolic traditions.

This flexibility positions the framework as a foundation for cross-cultural semantic composition, where diverse symbolic traditions can be interpreted as meaningful musical notation. Beyond technical adaptation, this approach supports artistic and educational applications, such as cultural heritage projects or immersive multimedia performances that link local symbolic vocabularies with dynamic sound and score structures. As a result, the system provides both a platform for technical experimentation and a space for cultural dialogue, with Chinese characters representing a significant and evolving branch of semantic-driven digital scoring.

4. CASE STUDY OF THE FRAMEWORK

In the prototype system developed for this research, **the moon** was chosen as the thematic focus.[1] Literary materials related to the moon in different languages were mapped into a semantic space to demonstrate the operability and cross-cultural applicability of the digital score framework.

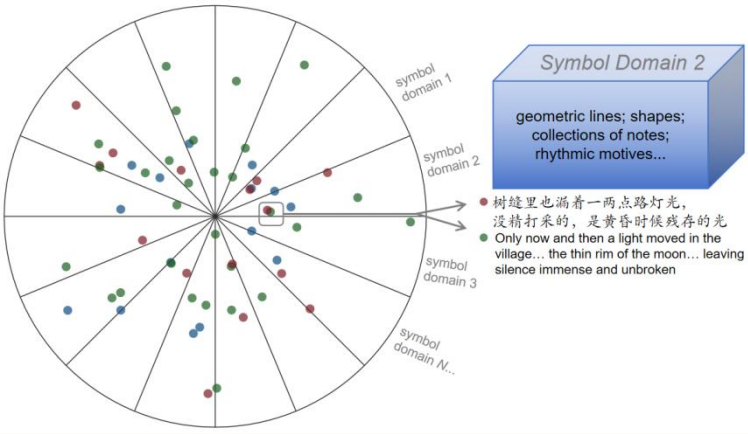

Figure 3. Semantic mapping of text fragments into symbolic domains for digital score representation

4.1. Data and Corpus

To ensure cultural and linguistic diversity, three categories of texts were selected:

- Chinese: Zhu Ziqing's modern prose-poem *Moonlight over the Lotus Pond* (荷塘月色).
- Latin: The medieval Latin poem *O Fortuna*(from *Carmina Burana*), in which the imagery of *Fortuna* as changeable as the moon serves as the connection.
- English: Moon-related passages from Virginia Woolf's novel *To the Lighthouse*.

[1]This case study isolates the notational/performative pathway; the sound-code pathway is implemented in the prototype but is not evaluated here to avoid potential confounds.

All of these texts contain imagery directly or metaphorically associated with the moon, forming resonances across different cultures and contexts.

4.2. Construction of the Semantic Space

The texts were segmented into their smallest semantic units: paragraphs for the Chinese material, stanzas for the Latin poem, and descriptive fragments for the English text. Each unit was represented as a semantic point, mapped into a high-dimensional semantic space and then projected into two dimensions through dimensionality reduction.

- Color: distinguishes language (Chinese = red, English = green, Latin = blue).
- Angle: divides the semantic space evenly into angular sectors, so that language points with different semantic orientations are clearly distributed.

In this way, composers and performers can visually observe how multilingual materials are distributed and clustered within the same semantic field.

4.3. Compositional Symbol Domains

The semantic space is evenly divided into *N* angular sectors, and for each sector a **compositional symbol domain** is defined.

Each sector functions as a compositional symbol domain, that is, a concept predetermined by the composer, such as a theme, pitch sequence, rhythmic pattern, or performance technique.

These symbols do not function as mechanical notes but as triggers for improvisation: when a performer encounters a point located within a given sector, they are required to improvise according to the musical content indicated by the corresponding symbol domain.

The semantic space thus becomes a conceptual map, establishing associations between natural language and musical symbols. This association makes full use of the properties of semantic space, breaks down barriers between languages, and ultimately establishes explicit mappings to musical symbols.

In terms of performance logic, improvisation is carried out according to the compositional symbol domains determined by the sequence of semantic points within the semantic space .

4.4. Summary of the Case

The moon-themed case study highlights the core features of the framework:

(1) A unified semantic space across languages: Chinese, English, and Latin texts are juxtaposed within the same space, allowing performers to intuitively perceive the convergence and divergence of cross-cultural imagery.

(2) Semantic similarity-driven spatial distribution: text units with similar meanings are distributed in adjacent regions, thereby creating a layered structure of semantics at the macro level.

(3) Composer's personalized symbol domains: provide performers with conceptual cues rather than fixed prescriptions.

(4) A performance logic combining sequence and symbol: the sequence is determined by the numbering of language points, while the symbols are created by the composer; improvisation emerges under the guidance of these symbols.

This case demonstrates the feasibility of the semantic-driven digital score and highlights its potential in uniting multilingual texts, musical composition, and performative interpretation.

5. CONCLUSION

The key insight of this study is that the proposed framework is not merely a mechanism for mapping between text and sound, but is gradually constructing **a new notational language** [7]. Unlike traditional scores that rely on abstract musical symbols, this language is rooted in semantic carriers (such as Chinese characters) and extends into multiple dimensions of visuality, sound, and performance.

This new language exhibits the core characteristics of a symbolic system:

(1)Lexical level: each symbol module mapped from a semantic point retains its cultural meaning while simultaneously encoding performance instructions.

(2)Syntactic level: the arrangement of these symbols along the temporal axis forms a kind of *grammar*, constrained by the semantic ordering of language points derived from cross-linguistic corpora.

(3)Pragmatic level: the same symbol may assume very different expressive roles in different poems and performance contexts. It is precisely this sensitivity to context — one of the hallmarks of pragmatics — that endows the symbolic language with elasticity, adaptability, and vitality, allowing it to continuously generate new meaning in use.

Within this framework, the composer is no longer merely a creator of musical materials, but becomes **a designer of language**. The role of composition is thus redefined: not simply the writing of fixed notes, but the curation of a symbol–semantic space in which performers and language models co-construct meaning. In this sense, composition is the invention and evolution of a digital notational language — one that is culturally grounded, semantically interpretable, and adaptable across languages and works. We evaluated the symbolic/performative path; systematic evaluation of the sound-code path is left as future work.

Acknowledgments

This paper is supported by the following projects:

Digital Scores – investigating the technological transformation of the music score
(DigiScore Grant agreement ID: 101002086)

Research Funding Program of the Central Conservatory of Music (Grant No. 23YB19).

6. REFERENCES

1. C. Vear, *The Digital Score: Musicianship, Creativity and Innovation*. Routledge, 2019.

2. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, vol. 1, 2019, pp. 4171–4186.

3. T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

4. E. Brown, "December 1952," in *Notations*, J. Cage and A. C. Childs, Eds. New York: Something Else Press, 1969, pp. 57–58.

5. C. Cardew, *Treatise*. London: Gallery Upstairs Press, 1967.

6. F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, "Language-agnostic BERT sentence embedding," in *Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, vol. 1, 2022, pp. 878–891.

7. N. Goodman, *Languages of Art: An Approach to a Theory of Symbols*. Indianapolis: Hackett Publishing, 1976.