

Poems as Musical Instruments: an AI Music Performance Interface Based on the Concepts of Chinese Qin Songs

Authors

Songruoyao Wu, Jiaxing Yu, Kejun Zhang, Wenqi Wu, Chongjun Zhong and Xiuqi Li

Information

Songruoyao Wu (researcher, designer), College of Computer Science and Technology, Zhejiang University, No.38 Yuquan Rd., Hangzhou, Zhejiang, 310027, China. Email: wsry@zju.edu.cn. ORCID: 0000-0002-0760-6289

Jiaxing Yu (researcher, developer), College of Computer Science and Technology, Zhejiang University, No.38 Yuquan Rd., Hangzhou, Zhejiang, 310027, China. Email: yujx@zju.edu.cn. ORCID: 0009-0008-2162-5110

Corresponding author: Kejun Zhang (educator), College of Computer Science and Technology, Zhejiang University, No.38 Yuquan Rd., Hangzhou, Zhejiang, 310027, China. Email: zhangkejun@zju.edu.cn. ORCID: 0000-0003-4592-1818

Wenqi Wu (researcher, designer), College of Computer Science and Technology, Zhejiang University, No.38 Yuquan Rd., Hangzhou, Zhejiang, 310027, China. Email: wenqi_wu@zju.edu.cn. ORCID: 0009-0001-8811-6360.

Chongjun Zhong (researcher, designer), College of Computer Science and Technology, Zhejiang University, No.38 Yuquan Rd., Hangzhou, Zhejiang, 310027, China. Email: zhong_chongjun@zju.edu.cn. ORCID: 0000-0002-8671-050X.

Xiuqi Li (student), College of Computer Science and Technology, Zhejiang University, No.38 Yuquan Rd., Hangzhou, Zhejiang, 310027, China. Email: 3210105694@zju.edu.cn. ORCID: 0009-0008-9217-0416.

Abstract: The authors have designed a new interface of musical expression called *Poem-Qin*, which reveals the relationship between text and music, supporting self-accompanied singing and cross-media musical practices. Drawing inspiration from Chinese qin songs and movable-

type printing, the authors creatively integrate Chinese characters into the interface, controlling music with tone inflections and linguistic rhythms. Moreover, they have developed an AI model for text-to-melody composing to assist in creating new music content. User and audience testing has demonstrated that it offers a positive experience and effective creative results.

1. The Dilemmas of Self-accompanied Singing Novices

With the rise of video social media, personal music performances have gained widespread distribution. Among these, self-accompanied singing, or providing instrumental music alongside one's own singing, has become very popular. Consequently, a surge of tutorial videos appears simultaneously. They explain instrument-playing techniques and try to make them easy to practice. However, their efforts may miss the point. The purpose of self-accompanied singing is to blend the voice with the instrument to enrich emotional expression. The playing techniques are merely a means to achieve this end.

To support this view, we studied *Chinese qin songs*, an ancient and simplified self-accompanied singing form. It is a significant expression of *Chinese guqin art*, which is a UNESCO intangible cultural heritage [1]. "Qin" is the instrument for playing the melody, while "song" denotes chanting and singing. According to historical records, the ancient Chinese literati believed that the purpose of music was to reveal one's inner self and to find kindred spirits rather than to display technical prowess [2].

The ancients' contemplation inspired our design (Fig.1). They believed that *qin songs* represented a harmonious symbiotic between different art forms (literature and music), different sounders (singer and instrument), and different expressions (concrete and abstract). It is instinctive, improvisational, and eclectic, immensely amplifying the performer's emotional and spiritual energy. Confucius often used it for cultural and educational activities. Academic research also proves that music can promote people's memory of words [3].

Like today's self-accompanied singing, the widespread adoption of *qin songs* faces obstacles regarding tools and skills. The instrument requires a complex crafting process to achieve the desired tonal quality. Moreover, learning to play it demands significant time and effort. These objective factors greatly hinder the average person's participation in this fascinating art form.

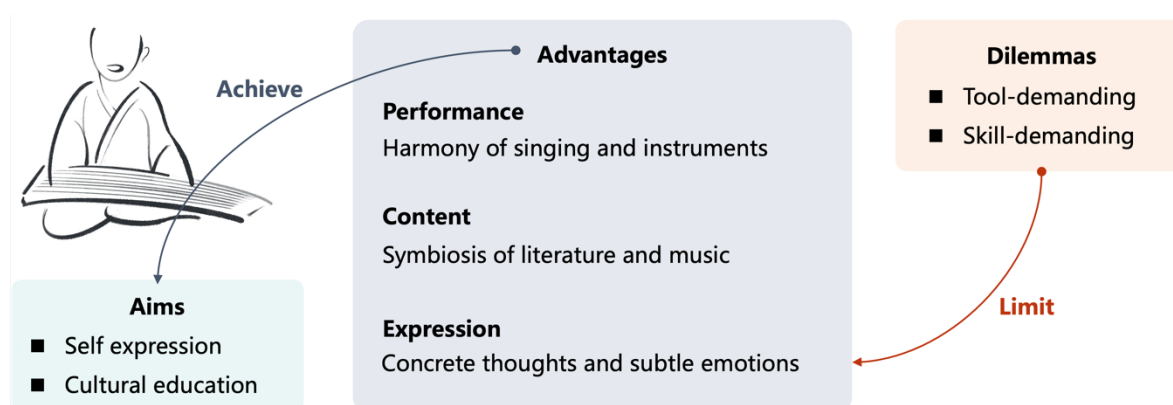


Fig. 1. The inspiration of *qin songs* to self-accompanied singing (©Songruoyao Wu)

2. New Technologies Supporting Self-Accompanied Singing

The above challenges may be resolved as new technologies are applied to art. For instance, digital audio and software technology offers more multi-sensory artistic experiences [4]. Smart devices and touchscreen technology provide more accessible virtual instruments and community resources [5,6]. With the globalization of digital music platforms, artificial intelligence (AI) increasingly assists in music production, experience, and performance. [7]. Therefore, we need a design medium to integrate multiple technical advantages.

New Interfaces for Musical Expression (NIME) incorporate cutting-edge technologies from various fields and consider their roles in music analysis, composition, interaction, and experience. NIME research originated from music performance practices, where performers use environmental and motion signals to control musical sounds [8]. Since 2002, the annual NIME international conference has been held, gathering researchers from fields such as music, design, electronics, information technology, psychology, and more.

With the advancement of digital technology, smart devices have become the new carrier of NIME. Among the NIME based on 2D touch interfaces, tablet-based products accounted for the largest share, reaching 29.4% [9]. These new NIMEs are loaded with various AI models. Surveys show that most people have a positive attitude toward AI music production tools [10].

Due to the complexity of musical practice, NIME has no uniform design standard. However, previous research and discussions identified some general principles [11]. Firstly, the goals of NIME design come from specific cultural and social backgrounds. It may be created to solve problems or build new meanings [12]. Secondly, NIME should support creative practice [13].

Their main goal is to provide exploratory art experiences to the public [14]. Finally, NIME exists in ecosystems and impacts multiple subjects [15]. Based on the above principles, we created an intelligent music interface to solve the difficulties of self-accompanied singing performances.

3. The *Poem-Qin* Musical Performance Interface

Based on the concept of *qin songs*, we have designed an interactive music interface based on intelligent mobile devices named *Poem-Qin*. It simplifies self-accompanied singing performance through a multi-faceted alignment of text and music [16].

3.1. An Instrument Composed of Chinese Characters

As the interface is aimed at novices, we make the text characters correspond to the musical notes to show the multimedia information of qin songs. We draw inspiration from another world's intangible cultural heritage originating in ancient China, the movable type printing technique [17]. Each Chinese character is designed as a module and is arranged into different textual works. The modules also serve as the keys of the musical instrument, containing visual, auditory, and textual information (Fig. 2a).

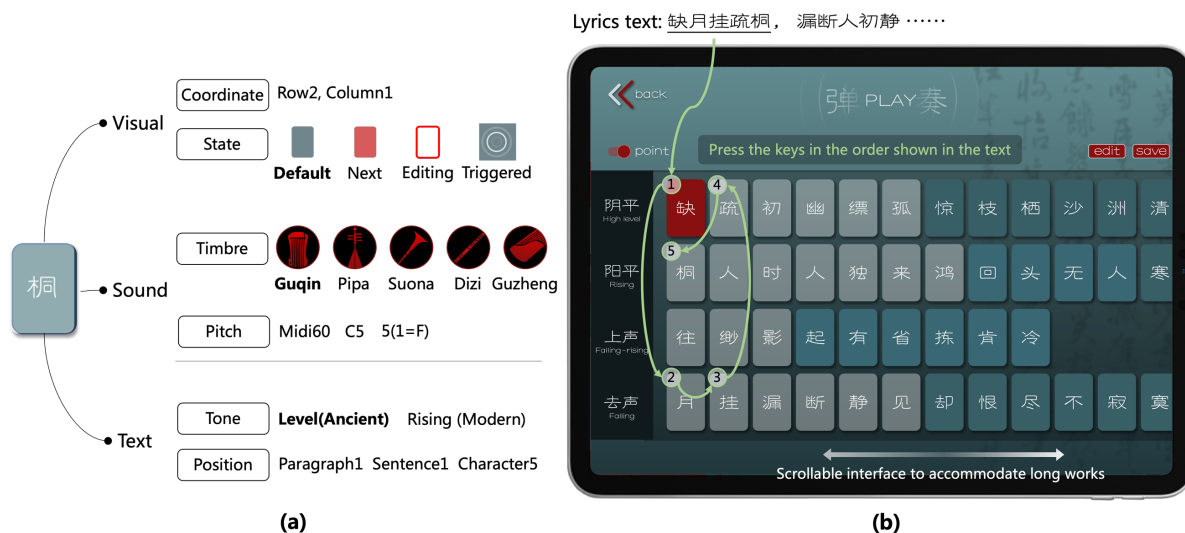


Fig. 2. The music performance interface: (a) a musical module (key); (b) the interactive process. (©Songruoyao Wu)

Considering the cultural functions of *qin songs*, we divided the interface into four lines

according to the tones of Chinese characters (Fig. 2b). Like the ancient guqin, each string (line) implies a pitch tendency. A scrollable view is used to increase the length of the instrument. Under this arrangement, users can find the next key within two to three columns and feel the relationship between tone and pitch (see the green line in Fig. 2b). The only condition for proficient performance is remembering the text of the ancient literary work. The appropriate difficulty encourages users to understand the connotations of literary works to achieve better performance, expression, and dissemination.

In the past, self-accompanied singing performers often multitask. For playing, they needed to remember the sequence of notes, the positions of the keys, and the correct beats. For singing, they needed to memorize the lyrics, to control their breath, and to manage their emotions. They had to be proficient in one or both to achieve perfect harmony.

By establishing multiple connections between text and music, our design turns playing and singing into one task. Music novices can perform self-accompanied singing based on the text. Literature enthusiasts can understand and memorize the literary works within the rhythm of the music. We hope people with different backgrounds will participate in this artistic practice together and offer their unique perspectives and creativity.

3.2. Relay Creation with Ancients or AI

NIME should offer more avenues for exploration and innovation. We hope users participate in the music creation process, showcasing their aesthetic preferences and expression. Therefore, *Poem-Qin* provides a song library and editing tools (Fig. 3).



Fig. 3. The interfaces: (a) the song library; (b) the edit process. (©Songruoyao Wu)

The library provides users with preset songs that will be updated continuously. It includes classic ancient *qin songs* and new pieces created by AI. Classical works contain thoughts of famous creators, which still guide today's songwriting. However, completely recorded *qin songs* are too few to satisfy users' experience needs. Therefore, we use an AI model to mimic the music patterns and create new melodies. The method of creation will be detailed in the next chapter.

In ancient times, the creation of *qin songs* was a relay process. The works had to be refined by different participants, such as literati, musicians, and singers. One piece would be used to express emotions in various contexts. Performers would impart their improvisation to the work, resulting in a series of derivative creations.

We aspire for classical and standardized works to guide users without constraining their creativity. Hence, *Poem-Qin* provides an editing mode that encourages users to express personalized artistic thoughts, building upon the foundation of the ancients and AI. The editing tool is embedded in the performance interface of each music piece, assisting users in viewing and modifying the pitch and timbre of each note module. It enables users to comprehend the melody and match it with text promptly. Ultimately, new compositions can be stored in the library for subsequent comparison, display, and exchange.

4. Building a Song Library with AI

AI melody generation requires substantial musical data or quantifiable creative theories. We adopted the latter due to the scarcity of fully documented *qin songs*. We collected 20 ancient *qin songs* and 20 modern (recomposed) pieces, analyzed the combination of text and melody within them, and summarized the music patterns of *qin songs* on three levels: individual characters, adjacent characters, and sentences.

4.1. The Relationship Between Individual Characters and Melody

"Tone" refers to the pitch variations of some languages' pronunciation [18]. The tones of Chinese Mandarin include level, rising, falling-rising, and falling. The official tones of ancient times included level, rising, short, and falling [19]. An online tool can be used to determine the tones of each character [20]. Figure 4 shows the correspondence between the eight tones and

the notes in ancient and modern qin song compositions. In over 3/4 of cases, one character corresponds to one note, consistent with the monosyllabic nature of Chinese. Less than 1/4 of the characters correspond to multiple notes. In this case, modern *qin songs* tend to choose a melody that coincides with the tone shape (indicated by the black-framed part), while ancient *qin songs* use ascending melodies.

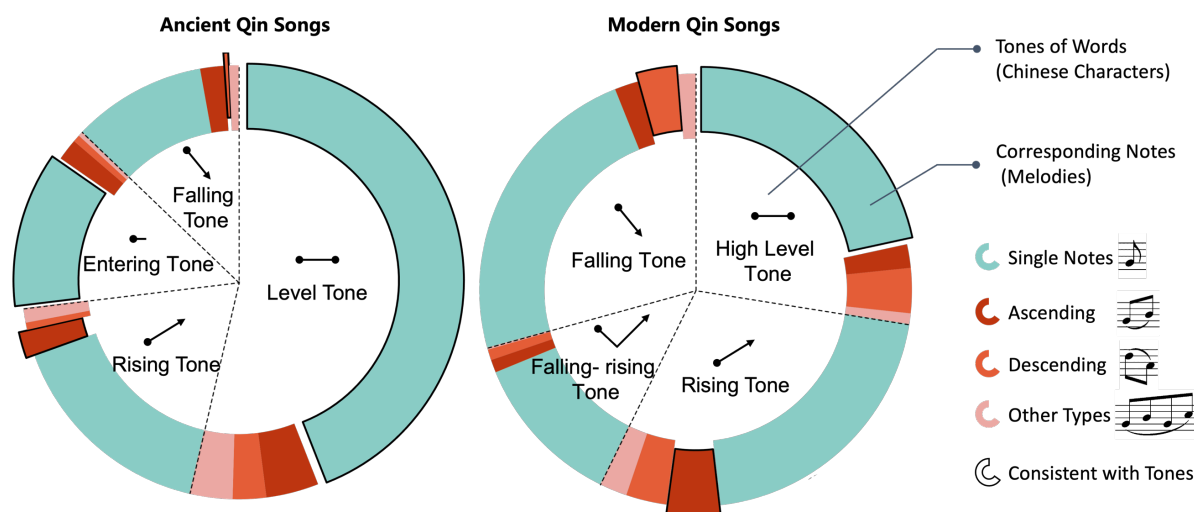


Fig. 4. The melodies (notes) correspond to individual characters (©Songruoyao Wu)

4.2. The Relationship Between Adjacent Characters and Melody

Ancient tones have 16 adjacent combinations, as do the modern ones. We analyzed each combination's pitch transitions with the calculation method illustrated in Fig. 5a. Figure 5b shows the statistical results. In modern *qin songs*, the combination of adjacent tones has a more noticeable impact on pitch transition. For example, the melody's pitch will significantly decrease when the falling-rising tone appears after the high-level tone. This is consistent with the pitch of the two tones in speech. In addition, modern *qin songs* have more obvious melodic fluctuations than ancient pieces, as the case in Fig. 5c.

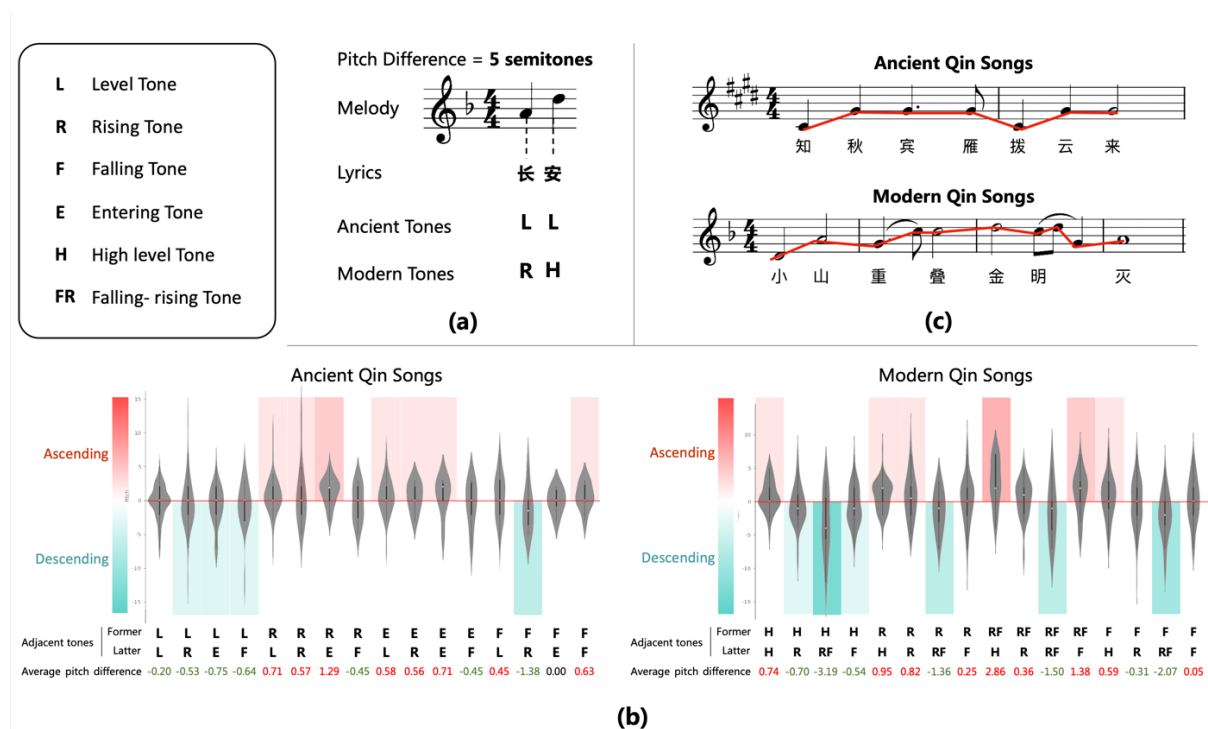


Fig. 5. The pitch transitions between adjacent characters: (a) calculation method, (b) statistical results, (c) melody cases. (©Songruoyao Wu)

4.3. The Relationship between Poem Structure and Melody

Structure dictates the number, length, and function of sentences within a text. In the Song Dynasty, a fixed text structure was known as a "Ci Pai" (tune name), which was also associated with a specific musical melody [21]. The lyrics of qin songs typically comprise one or two paragraphs. Each paragraph contains four sentences, which serve to establish the theme, continue the story, introduce variations, and summarize emotions. Correspondingly, its melody will exhibit repetition or partial repetition. Based on the collected songs, we have identified eight patterns of repetition, as shown in Fig. 6.

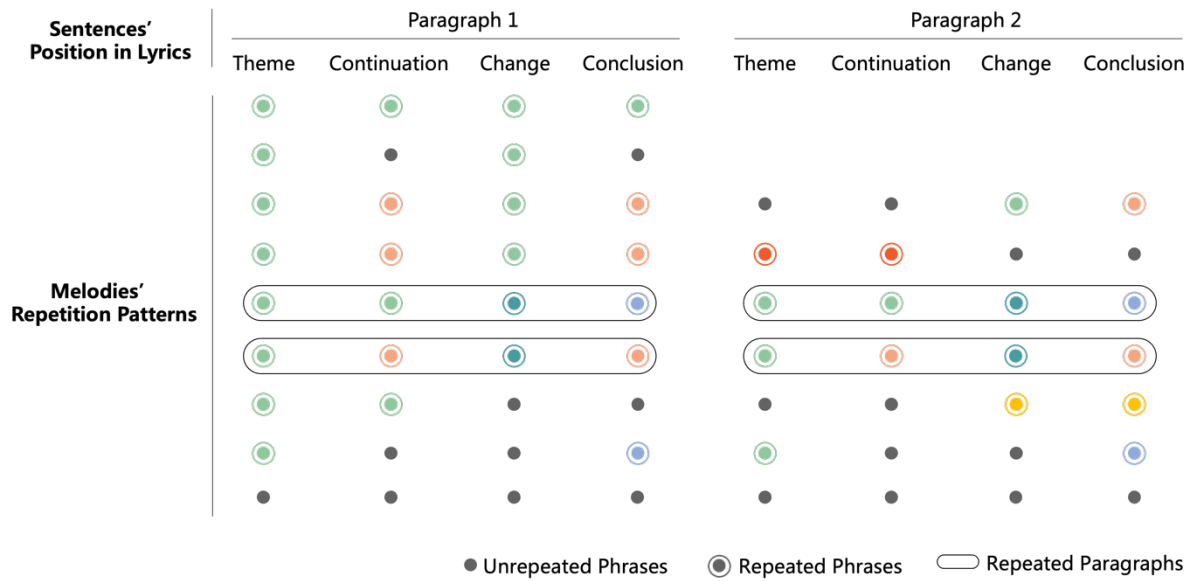


Fig. 6. The melodies' repetition patterns (points of the same color, except black, indicate repeated or similar musical phrases) (©Songruoyao Wu)

4.4. Text-Driven Melody Generation

We introduced AI models to generate melodies to replenish the music library continuously. The patterns mentioned above are utilized to constrain the model's generative process. AI models like TeleMelody can create melodies based on lyrical text. Still, they lack the necessary attention to the relationship between the two media forms, which leads to disharmony and misunderstanding [22]. Based on our previous research [23], we apply two sets of qin song patterns (ancient and modern) as constraints to the model's output (Fig. 7). It offers "suggestions" in pitch and repetition, making the generated melody suited to the text.

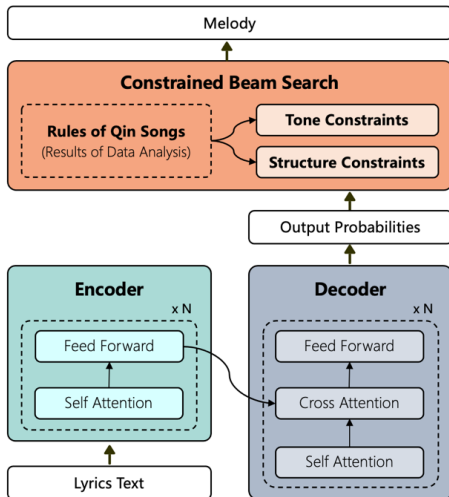


Fig. 7. The AI melody generation model (the orange part shows our optimization work)
(©Songruoyao Wu and Jiaxing Yu)

5. User Experience and Feedback

5.1. Test Process

We conducted experiments with performers (n=20) and audiences (n=30). In the performers' experiment, we tested the user experience and cultural dissemination capability of *Poem-Qin*. We encouraged participants to share their feelings and suggestions beyond our design objectives. In the audience experiment, we used *Poem-Qin* to create songs in ancient and modern styles based on the lyrics of existing *qin songs*. The audiences reported their subjective evaluations of the original and the new works without knowing the composers. All songs were produced and presented in the same way.

5.2. Feedback from Performers

The text-based interface provides many meaningful cues that help users discover tricks in playing. The novices believe that "the musical keys with Chinese characters alleviate some memory burden and reduce difficulty". The interaction pattern fits the logic of the language and thus becomes simpler. "Self-accompanied singing usually requires multitasking. However, this interface unifies playing and singing, allowing me to experience the music more fully." They also found that playing *Poem-Qin* requires familiarity with the corresponding literary works and suggested the interface should include more explicit cues to help children or cross-cultural users.

During the creation process, the AI model becomes a collaborator. Although its output lacks explanation, users would actively adjust their creative thinking based on the "hints" it provides. "In my previous creation experiences, I had not paid attention to the role of tone in music, but this interface reminded me of that." We are pleased to see the inspiration AI brings to human creators. These influences may be lasting, guiding users to contemplate new frameworks for understanding and creating music.

At the end of the experiment, we introduced users to the origins of *qin songs*. Despite the interface being utterly different from traditional guqin instruments, most users still felt that "its connection to *qin songs* is very evident". The *qin song* rules embedded in the interface are

subtle, and users need to discover them through repeated practice. "When I don't press the keys in the prescribed order, I feel the pitch difference between the rows." Like Easter eggs, *Poem-Qin* sparks further exploration interest.

5.3. Feedback from Audiences

We utilized a seven-point Likert scale to collect audience evaluations. Ancient-style *qin songs* created with *Poem-Qin* achieved evaluations that were close to the original works (Table 1). In creating modern *qin songs*, *Poem-Qin* significantly improves the melody's harmony and aesthetic evaluation (Table 2). This indicates that audiences can discern whether the text and melody of *qin songs* are well-matched. The harmony between them can reduce misunderstandings, increase memory points, and enhance aesthetic evaluation.

6. Conclusion and Prospect

Unlike music generation tools that treat lyrics merely as input, *Poem-Qin* demonstrates the symbiosis of literature and music from three dimensions: content creation, interface design, and interactive experience. It integrates the concept of *qin songs* with knowledge from different disciplines, representing a cross-disciplinary practice in art inheritance and innovation. Our main contributions include:

- (1) We designed a NIME for self-accompanied singing. Experiments have proven that it has good cultural dissemination and art education functions.
- (2) We summarized the patterns in ancient and modern *qin songs* from the levels of characters, words, and sentences, proposing a quantifiable theory of melody creation.
- (3) We developed an automated composition method for qin song music by integrating AI technology, providing a new practical approach for text-to-melody creation.

Our research indicates that traditional art encompasses concise and intriguing design concepts. With the support of technology, they can demonstrate incredible value. Moreover, we have serendipitously discovered a positive collaboration between humans and AI in music composition. AI excels in analysis and induction, which can alleviate the computational burden on humans and support more experimental exploration.

We have collected suggestions from users and audiences, including cross-media interaction, style fusion, dynamic feedback, and multi-person collaboration. This has inspired us to think about other ways of musical expression. Therefore, we are considering two tasks to be completed in the future: (1) Establish a multimedia traditional music database to explore richer artistic characteristics and expression modes. (2) Design cross-media tools for human-AI collaborative creation to promote artistic innovation and exploration. We believe that new technologies and design methods can empower humanity to revive the beauty of the past and pioneer the beauty of the future.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62272409); the Key R&D Program of Zhejiang Province (No. 2022C03126); and the Ministry of Culture and Tourism (No. 2022DMKLB001).

References and Notes

1. Chinese Academy of Arts, "Guqin and Its Music," UNESCO Intangible Cultural Heritage <ich.unesco.org/en/RL/guqin-and-its-music-00061> accessed 14 December 2023.
2. Z. Qian, and C. Schoenberger, "Music, Morality, and Genre in Tang Poetry," *Journal of Chinese Literature and Culture* **5**, No.1, 66--94 (2018).
3. K. L. Winters and S. Griffin, "Singing is a Celebration of Language: Using Music to Enhance Young Children's Vocabularies," *Language and Literacy* **16**, No.3, 78--91 (2014).
4. Y. Yang et al., "Interactive Moiré Patterns Reflecting on the Traditional Nanjing Baiju." *Leonardo* **56**, No.1, 10-16 (2023).
5. G. Wang, "Some Principles of Visual Design for Computer Music." *Leonardo Music Journal* **26** (2016) pp. 14-19.
6. O. Vallis and A. Kapur, "Community-based design: The Democratization of Musical Interface Construction." *Leonardo Music Journal* **21** (2011) pp. 29-34.
7. N. Bryan-Kinns, Z. Li and X. Sun, "On Digital Platforms and AI for Music in the UK and China." Proceedings of the International Conference on New Interfaces for Musical Expression

(2020) pp. 357–360.

8. M. Gurevich, "Diversity in NIME Research Practices." *Leonardo* **49**, No.1, 80-81(2016).
9. D. Schwarz, W. Y. Liu and F. Bevilacqua, "A Survey on the Use of 2D Touch Interfaces for Musical Expression," *Proceedings of the International Conference on New Interfaces for Musical Expression* (2020) pp. 196--201.
10. E. Frid and A. Ilsar, "Reimagining (Accessible) Digital Musical Instruments: A Survey on Electronic Music-Making Tools," *NIME 2021*.
11. J. Barbosa et al., "What Does Evaluation Mean for the NIME Community?" *Proceedings of the International Conference on New Interfaces for Musical Expression* (2015) pp.156--161.
12. L. Dahl, "Designing New Musical Interfaces as Research: What's the Problem?" *Leonardo* **49**, No. 1, 76--77 (2016).
13. A. Johnston, "Opportunities for Practice-Based Research in Musical Instrument Design," *Leonardo* **49**, No. 1, 82–83 (2016).
14. Holland Simon et al. ed., *New Directions in Music and Human-Computer Interaction*. (Berlin: Springer, 2019) pp. 187--197.
15. M. Rodger et al., "What Makes a Good Musical Instrument? A Matter of Processes, Ecologies and Specificities," *Proceedings of the International Conference on New Interfaces for Musical Expression* (2020) pp. 405--410.
16. The available demo is at <nextlab-zju.github.io/poemqin/>, accessed 8 June 2024.
17. Culture Ministry of China, "Wooden Movable-Type Printing of China," UNESCO Intangible Cultural Heritage <ich.unesco.org/en/USL/wooden-movable-type-printing-of-china-00322> accessed 18 March 2024
18. Yip Moira, *Tone* (Cambridge: Cambridge University Press, 2002.) pp. 1--4.
19. C. C. Cheng, "A Quantitative Study of Chinese Tones." *Journal of Chinese Linguistics* **1**, No. 1, 93–110 (1973).

20. The pronunciation is got at <ma.voicedic.com/> accessed 14 December 2023
21. B. Wang et al., "Automatic Recognition of Tune Names of Song Ci-Poetry," *2018 International Conference on Asian Language Processing* (2018) pp. 189--192.
22. Z. Q. Ju et al., "TeleMelody: Lyric-to-Melody Generation with a Template-Based Two-Stage Method," *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (2022) pp. 5426--5437.
23. C. Zhang et al., "ReLyMe: Improving Lyric-to-Melody Generation by Incorporating Lyric-Melody Relationships," *Proceedings of the 30th ACM International Conference on Multimedia* (2022) pp. 1047--1056.

Bibliographical Information

SONGRUOYAO WU is a PhD candidate in Artificial Intelligence at Zhejiang University. Her research focuses on artificial intelligence-assisted art creation and design.

JIAXING YU is a PhD candidate at Zhejiang University. His research interests include natural language processing, music information retrieval, and multimedia.

KEJUN ZHANG is a professor and head of the Department of Industrial Design at the College of Science and Technology, Zhejiang University. His research focuses on the intersection of art, science, and design. He is committed to using innovative technology to enhance the beauty of traditional art. More details: <https://next.zju.edu.cn>.

WENQI WU is a PhD candidate in design at Zhejiang University. She is an enthusiast of traditional art and is currently working on using AI to enhance artistic preservation and innovation in intangible cultural heritage.

CHONGJUN ZHONG is a PhD candidate in design at Zhejiang University. Her research interests include human-computer interaction, digital art, and music co-creation.

XIUQI LI is an undergraduate student in design at Zhejiang University. She is strongly interested in traditional culture, excels in interface design, and is passionate about AI-assisted art creation.

Tables and charts

Ancient Qin-Songs (Original Work – Recomposed with *Poem-Qin*)

Indicator	t	df	p	Mean Difference	SE Difference	95% CI for Mean Difference	
						Lower	Upper
Listenability	-1.769	58	0.082	-0.533	0.301	-1.137	0.070
Harmony	-0.467	58	0.642	-0.150	0.321	-0.793	0.493

Table 1. Results of independent t-test. Comparison between the original works (ancient *qin songs*) and the works recomposed with *Poem-Qin*

Modern Qin Songs (Original Work – Recomposed with *Poem-Qin*)

Indicator	t	df	p	Mean Difference	SE Difference	95% CI for Mean Difference	
						Lower	Upper
Listenability	-1.968	58	0.027 *	-0.500	0.254	-1.009	0.009
Harmony	-2.235	58	0.015 *	-0.550	0.246	-1.043	-0.057

Table 2. Results of independent t-test: the evaluation of original works (modern *qin songs*) is less than the works recomposed with *Poem-Qin* (statistically significant results reported as $p < 0.05$ *)