Human-Computer Interaction & Emerging Technologies, Vol. 233, 2026, 98–105
https://doi.org/10.54941/ahfe1007508

Designing a Rhythmic AR Interaction for Auditory-Oriented Heritage: A Preliminary Case Study at Guqintai

Yitong Chen and Miaoxin Zhang

School of Art and Design, Wuhan University of Technology, Wuhan 430070, China

ABSTRACT

As digital cultural tourism increasingly shifts from one-way presentation to participatory interaction, augmented reality (AR) technology has been progressively applied in cultural heritage contexts. However, most existing AR cultural tourism applications remain limited to overlays of text, images, or animations, leaving users largely in a passive viewing role. Systematic research on how embodied interaction influences users' comprehensive engagement and psychological perception, as well as on perceptual differences across various types of heritage sites, remains scarce. This study takes the 'Zhi Yin' culture carried by Wuhan's Guqintai as a case study. We developed a rhythm-driven AR interaction prototype that transforms the classical narrative of Bo Ya and Zhong Ziqi into actionable interactive experiences, guiding visitors to perceive narrative continuity and develop a sense of situated engagement during their on-site visit. A small-scale field user test was conducted, collecting observational notes and interview data to evaluate system usability, immersion, and cultural understanding. Results indicate that rhythm-driven interaction can markedly improve engagement in auditory-oriented heritage settings: most participants quickly learned the rhythm-trigger logic, and the sound-motion synchronization mechanism effectively increased attention and emotional involvement. Furthermore, gesture amplitude was positively correlated with immersion across different rhythm stages, with large gestures (e.g., in punching interaction) eliciting higher engagement and enjoyment than smaller gestures (e.g., in conducting interaction). In summary, this study provides preliminary practical and theoretical insights for AR interaction design in auditory-oriented heritage contexts. It demonstrates the potential of rhythm-driven interaction to enhance engagement, facilitate cultural understanding, and evoke emotional resonance, laying the foundation for future large-sample studies and adaptive, data-driven optimization research.

Keywords: Augmented reality, Rhythm-based interaction, Auditory-oriented heritage, Digital cultural tourism, Embodied experience

INTRODUCTION

As digital technologies are gradually enabling more participatory forms of engagement, heritage tourism is no longer limited to one-way information delivery. Recent studies suggest that AR has progressed beyond simple "visual augmentation" to support co-created heritage and sensory experiences (Shafiee Roodposhti and Esmaeelbeigi, 2024), cultivating cultural resonance through a sense of intertemporal connectedness (Tang and Zhou, 2025). By

Received March 5, 2026; Revised April 11, 2026; Accepted April 25, 2026; Available online July 20, 2026

adding text overlays or 3D models, most existing AR applications are still used primarily to enrich exhibitions (Bekele et al., 2018). While this improves visual presentation, interaction remains limited. Recent studies emphasize that the integration of interactive gaming and narrative mechanisms is vital for bolstering heritage appeal and immersive quality (Liu, 2024; Liu et al. 2025). At Guqintai, where music and storytelling constitute the core cultural experience, visual displays alone are insufficient to communicate its rhythmic and emotional depth. Consequently, users often persist in a state of "passive observation" rather than achieving "embodied engagement."

Based on embodied cognition theories proposed by Dourish (2001) and Barsalou (2008), meaning arises through the interaction between bodily perception and the surrounding physical environment. At Guqintai, the "Zhi Yin" culture is understood through more than listening. It also involves physical participation, such as gesturing or punching.

In response, this study takes the "Zhi Yin" culture of the Guqintai in Wuhan as its context and develops a rhythm-driven multimodal AR prototype using Rokid AR glasses. The Stimulus–Organism–Response (SOR) framework is adopted to clarify the relationship between rhythmic stimulus and user responses, translating auditory-oriented heritage into an immersive and intertemporal experience through a rhythm-driven AR system architecture and a four-stage user experience path. Through the above design and empirical exploration, this study validated the feasibility of a rhythm-driven multimodal interaction framework within auditory-oriented cultural heritage contexts, and provided preliminary empirical evidence for the coupling relationship among embodied participation, rhythmic structure, and narrative mechanisms.

RHYTHM-DRIVEN AR FRAMEWORK

Cultural Context and Design Objectives

The Guqintai is located in Wuhan and is rooted in the well-known story of "High Mountains and Flowing Water," in which Bo Ya plays the qin and Zhong Ziqi understands his music. Its heritage value lies not in visual scenery, but in the emotional resonance carried by music. As an auditory-oriented site, its cultural meaning is difficult to grasp through brief, sightseeing-style visits alone.

Therefore, this study aims to move beyond the common visual focus in heritage design. Using the Guqintai as a case, we develop a multimodal AR interaction mechanism that encourages visitors to engage through rhythmic interaction and become active participants in cultural meaning-making.

Multimodal Interaction Mechanism

This study utilizes Dongas et al.'s (2023) four-dimensional model to make rhythm the main driver of interaction, applying sensorimotor synchronization (SMS) (Repp and Su, 2013). Research also shows that rhythmic cues can synchronize brain activity (Nozaradan, 2014). Within the "Zhi Yin" cultural context, rhythm transcends musical attributes to serve as Bo Ya's vehicle for emotional articulation. By transforming rhythm into actionable AR tasks, the system facilitates a cultural journey through the tight coupling of action and perception.

Rhythmic interaction is embedded within a narrative sequence— "Mutual Understanding," "String Breaking," and "String Renewal"—and supported by gamified incentives (Aebli, 2019). In the SOR framework, rhythm acts as the Stimulus (S). The user's bodily involvement and emotional reaction constitute the Organism (O). The resulting immersion, along with measurable outcomes such as user engagement, represent the Response (R).

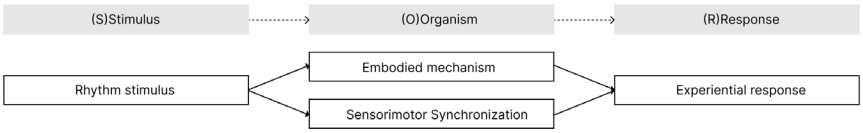

Figure 1: Stimulus-organism-response (SOR) framework applied to the AR rhythm interaction (self-drawn).

Prototype Development

To examine the feasibility of the rhythm-driven multimodal interaction framework in an auditory-oriented cultural heritage context, our team completed hardware capability integration and environmental adaptation based on the Rokid AR ecosystem, and developed the augmented reality interactive prototype using the Unity engine. The overall interface and 3D model design utilize a predominantly teal-green color scheme, referencing the teal tones of the Guqintai architecture to enhance scene coherence. At the same time, to avoid confusion between virtual and real elements, the traditional instruments, which are mainly red and black in reality, are rendered in bluish tones in the conducting mini-game prototype. This approach maintains cultural recognizability while clearly separating virtual elements from the physical environment. Interface layouts and interactive visual elements were initially designed and prototyped in Figma, then integrated into Unity for dynamic interaction.

At the structural level (Figure 2), the prototype adopts a stage-based game framework that closely links the real-world setting of Guqintai with interactive gameplay segments.

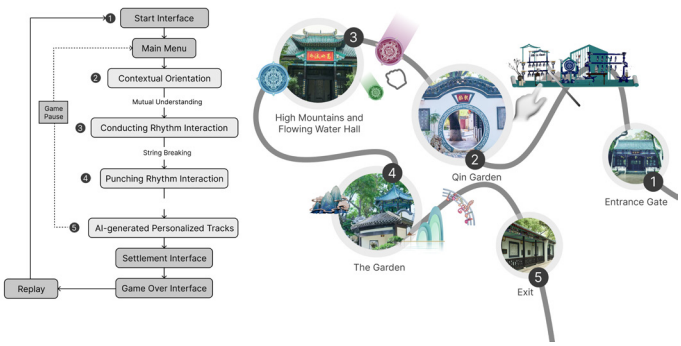

Figure 2: Stage-based game framework and spatial route mapping (self-drawn).

The overall design follows a "rhythm-action-narrative" coupling logic. Grounded in the Zhi Yin (soulmate) legend, the story was translated into a three-phase experiential path— "Mutual Understanding," "String Breaking," and "String Renewal." At the mechanism level, the prototype is organized according to both the actual visiting sequence of Guqintai—Entrance Gate, Qin Garden, High Mountains and Flowing Water Hall, and the Garden—and the narrative progression of "Bo Ya meeting Zhong Ziqi – Ziqi's passing – breaking the strings to express grief – renewing the strings to convey remembrance." The entire experience is thus divided into four stages, each corresponding to a specific spatial node.

This progressive structure, shaped jointly by spatial movement, narrative development, and increasing rhythmic intensity, aligns the strength of users' embodied participation with the emotional tension of the story, enabling a synchronized evolution of narrative logic and interaction mechanics.

Such a design not only achieves a high degree of alignment between the game framework and the physical site, but also provides a transferable reference for applying rhythm-driven interaction in other auditory-oriented cultural heritage contexts.

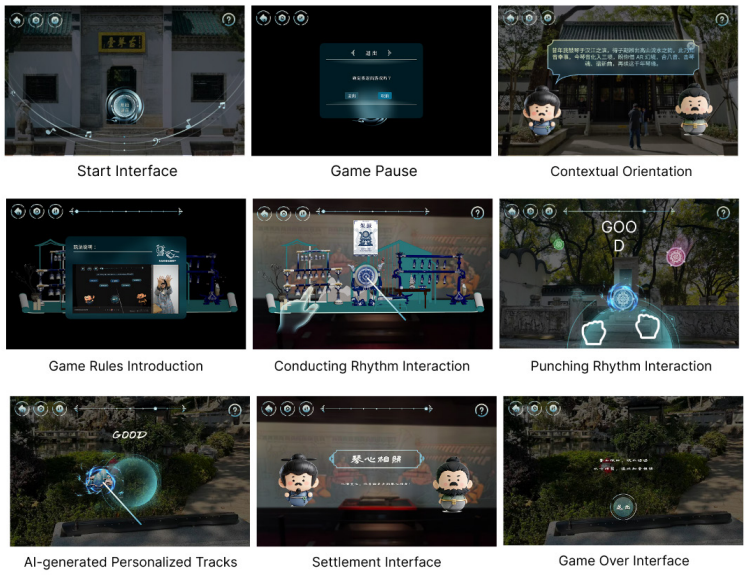

Figure 3: AR prototype interface screens (3 × 3 Layout).

System Architecture and User Experience Path

The system is built on a three-layer architecture to support rhythm-driven, multimodal interaction (Figure 4). The perception layer collects head movement, gestures, and spatial data via the Rokid AR glasses to calculate the user's rhythm synchronization. The processing layer maps this synchronization to the narrative progression, while the output layer delivers multimodal feedback through spatial audio and visual cues, keeping rhythm at the center of the experience.

We designed the user experience in four stages (Figure 5). It starts with contextual introduction, followed by the "Mutual Understanding" stage, where users conduct a rhythm game to explore the "Eight Sounds of Chu." Next, the "String Breaking" stage triggers emotional release through striking actions. Finally, the "String Renewal" stage uses AI to generate personalized music, creating each user's unique "Sound of Zhi Yin." As the narrative unfolds, the experience rises to an emotional peak and guides users toward a deeper understanding of the culture.

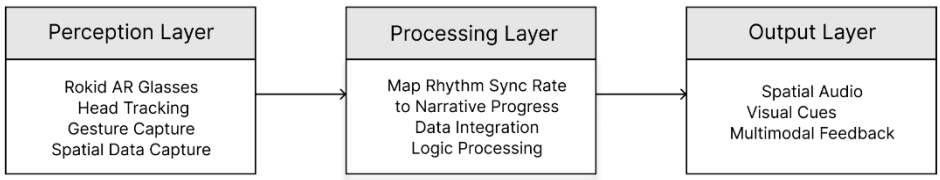

Figure 4: System architecture of the rhythm-driven AR interaction (self-drawn).

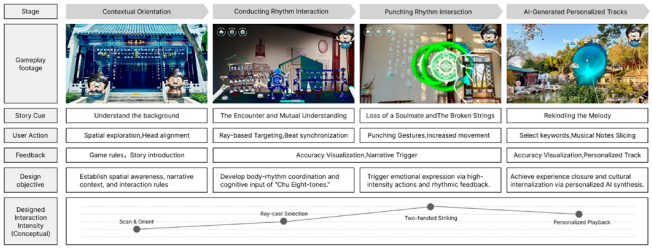

Figure 5: User experience path across four interaction stages (self-drawn).

User Evaluation

To test the feasibility of our rhythm-driven AR interaction in a hearing-focused cultural heritage context, our team conducted an exploratory field study at Guqintai (Figure 6). We recruited 12 participants, including tourists and university students. Most participants were familiar with "Zhi Yin" culture but had limited knowledge of Chu music and were using AR glasses for the first time.

Participants received a brief three-minute introduction before completing the four-stage rhythm interaction tasks, with each session lasting 20–30 minutes. On average, each participant completed a 25-minute narrative experience across all four stages. We collected usability and immersion ratings via questionnaires and conducted semi-structured interviews based on video playback, then applied thematic coding to capture changes in participants' perception of "Zhi Yin" culture.

The results showed that most participants quickly grasped the rhythm-trigger logic. The system was generally usable, and the action–sound synchronization appeared to enhance attention and engagement. During complex rhythm stages, recognition stability affected interaction continuity.

For example, target alignment in the conducting game was sometimes offset by spatial depth, and gesture recognition in the striking game occasionally lagged. These findings indicate that system calibration and recognition accuracy still require improvement.

Overall, the rhythm-driven interaction demonstrated potential to enhance immersion and emotional engagement in real-world heritage settings, though further testing with larger samples is needed.

Figure 6: Participants performing AR interaction.

DISCUSSION

The experimental results indicate that the degree of embodied participation has a significant positive effect on the perceptual experience of auditory-oriented cultural heritage (Suo et al., 2026). The Stimulus–Organism–Response (SOR) framework was employed as the evaluation standard to analyze the causal link between rhythmic stimuli and user engagement. During the four-stage interaction, users gradually moved more and became more emotionally engaged, suggesting a link between bodily participation and immersion. The punching game, which required larger movements, boosted both enjoyment and engagement more than the conducting rhythm game. By matching their gestures and movements to the rhythm, users physically synchronized with the music and emotionally resonated with the story of Bo Ya mourning his friend, which strengthened the overall immersive experience.

During the "String Renewal" stage, the system incorporated a generative AI personalization mechanism (Martusciello et al., 2025), allowing users to generate their own "Zhi Yin Sound" based on selected keywords. This feature exemplifies the central role of gamification in enhancing participation motivation and experiential benefits (Pasca et al., 2021). The mechanism not only increased participants' emotional engagement but also deepened cultural understanding and memory, demonstrating that personalized, rhythm-driven interactions can effectively integrate music, movement, and narrative for cultural translation.

The study's limitations include a small sample size of only 12 participants, potential interference from lighting, spatial layout, and ambient noise affecting motion recognition accuracy, and gesture recognition delays causing fluctuations in rhythm synchronization during complex stages.

Future research could incorporate physiological measures (e.g., heart rate, EEG) and cognitive load assessments to quantify the relationship between rhythm synchronization, immersion, and emotional resonance. Additionally, behavioral data such as movement amplitude and rhythm synchronization rates could serve as input parameters for AI systems to construct more embodied, adaptive cultural experience models, enabling dynamic adjustment of interaction difficulty and personalized content generation.

CONCLUSION

This study focuses on the "Zhi Yin" legend of Wuhan's Guqintai as its research object. Addressing the characteristics of auditory-oriented cultural heritage, we developed a rhythm-driven Augmented Reality (AR) prototype based on the narrative sequence of "Mutual Understanding—String Breaking—String Renewal" and conducted a preliminary validation of its interaction mechanism. Results from a small-scale field experiment indicate that, compared with conventional visually oriented AR tours, integrating narrative with sound, motion, and spatial cues encourages visitors to shift from passive viewing to active participation. Theoretically, this work provides a reference for AR design in auditory-oriented heritage; methodologically, it establishes a scalable rhythm-driven interaction framework, laying the groundwork for future adaptive and data-informed enhancements.

ACKNOWLEDGMENT

We would like to express our gratitude to our supervisor for their meticulous guidance and support in this study, as well as to the team members for their hard work. We would also like to thank the management of Guqintai and all the volunteers who participated in the experiment. Without their help, this study would not have been able to be completed smoothly.

REFERENCES

Aebli, A. (2019). Tourists' motives for gamified technology use. Annals of Tourism Research, 78, 102753. https://doi.org/10.1016/j.annals.2019.102753

Barsalou, L. W. (2008). Grounded cognition, Annual Review of Psychology, 59, 617–645.

Bekele, M. K., Pierdicca, R., Frontoni, E., Malinverni, E. S. and Gain, J. (2018). A survey of augmented, virtual, and mixed reality for cultural heritage. Journal on Computing and Cultural Heritage, 11(2), 1–36. https://doi.org/10.1145/3145534

Dongas, R. and Grace, K. (2023). Designing to Leverage Presence in VR Rhythm Games. Multimodal Technologies and Interaction, 7(2), 18.

Dourish, P. (2001). Where the Action Is: The Foundations of Embodied Interaction. MIT Press. https://doi.org/10.7551/mitpress/7221.001.0001

Liu, K. (2024). Application of Augmented Reality in Archive Cultural Heritage. Shanxi Archives, (6), 132–134.

Liu, X., Li, K., Geng, G. and Zhou, M. (2025). Review on Virtual Reality Technology in Cultural Heritage Protection. Journal of Computer-Aided Design & Computer Graphics, 37(4), 545–556.

Martusciello, F., Muccini, H. and Bucchiarone, A. (2025). A Reference Architecture for Gamified Cultural Heritage Applications Leveraging Generative AI and Augmented Reality. arXiv preprint arXiv:2506.04090.

Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1635), 20130393.

Pasca, M. G., Renzi, M. F., Di Pietro, L. and Guglielmetti Mugion, R. (2021). Gamification in tourism and hospitality research in the era of digital platforms: a systematic literature review. Journal of Service Theory and Practice, 31(5), 691–737.

Repp, B. H. and Su, Y. H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). Psychonomic Bulletin & Review, 20, 403–452.

Shafiee Roodposhti, M. and Esmaeelbeigi, F. (2024). Viewpoints on AR and VR in heritage tourism. Digital Applications in Archaeology and Cultural Heritage, 33, e00333.

Suo, X., Yin, B. and Feng, X. (2026). Augmented reality and embodied learning: Effects of embodiment degrees on students' learning achievement, cognitive load, and technology acceptance. Frontiers in Psychology, 16, 1712261.

Tang, Y. and Zhou, Q. (2025). Inspired by intertemporal connections: Using AR technology to enhance visitor satisfaction in historical museums. Tourism Management, 108, 105096. https://doi.org/10.1016/j.tourman.2024.105096