Vol.7 Iss.4 2026 Communication & Education Review
https://doi.org/10.37420/j.cer.2026.050

# A Study on the Application Strategies of Multimodal Visual Expression in the Interface Design of Digital Museum Guidance Systems

**Jia Wang*, Yi Yin**
**College of Arts, Jing Chu University of Technology, Hubei, China**
**Corresponding author e-mail: wangj920606@163.com**

**Abstract:**
*As museums are shifting from digital archiving to experiential guidance, existing digital guidance interfaces suffer from visual overload, modal isolation and fragmented cognition, which weaken immersion and emotional resonance. This study uses qualitative analysis and case studies of representative digital museum projects in China and abroad, proposes a triangular structural model of visual-auditory-tactile collaboration in digital museum guidance interface design. On this basis, the study summarizes key design principles and scenario-based application strategies for multimodal guidance interfaces, provides a theoretical framework and practical reference for optimizing information transmission, enhancing immersive experience and improving the digital communication of cultural heritage.*

**Keywords:**
*Digital Museum; Multimodal Interaction; Guidance Interface Design; Sensory Narrative*

## 1.Introduction

With the rapid advance of global digitalization, traditional museums are shifting from simple digital archiving toward experiential guidance, requiring new interface models that respond to the demand for immersive experiences[1]. This means that audiences no longer expect merely passive visual viewing, but increasingly seek forms of engagement that integrate perception, interaction and emotional resonance[2]. However, most existing guidance interfaces suffer from modal isolation, excessive dependence on visual presentation, and fragmented information organization[3]. Such problems not only increase cognitive load, but also weaken the immersive quality. Against this background, this study uses qualitative research and case analysis, examines representative digital museum projects in China and abroad, identifying common patterns of multimodal coupling and interface organization[4][5]. On this basis, it proposes a stable and efficient triangular structural model of visual-auditory-tactile collaboration, aiming to optimize the logic of information flow, reduce cognitive fragmentation and support the construction of integrated sensory narratives.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

## 2. Theoretical Review of Multimodal Perception in Digital Museum Guidance

### *2.1 Multimodal Interaction and Embodied Cognition*

Multimodal interaction refers to a comprehensive process of information exchange between users and digital systems through multiple perceptual channels, including vision, auditory, tactile, and behavioral operations. Furthermore, multimodal interaction moves beyond the traditional information-delivery model dominated by single visual channel and constructs a more multidimensional system of information expression through the coordination of multiple sensory modalities. From the perspective of cognitive science, this process was explained by the theory of embodied cognition, holding that cognition is not confined to internal information processing in the brain, but is instead a dynamic process jointly generated through bodily perception, active engagement, and interaction with the environment[1]. In this sense, multimodal interaction was essentially a coupling mechanism of "perception–action–environment," which reconstructed the cognitive pathway of information understanding through the circulation of multisensory input and behavioral feedback[2][3].

Through semantic complementarity and temporal coordination among different sensory channels, information was integrated and expressed across multiple dimensions, thereby reducing the cognitive burden placed on any single channel and enhancing the depth and stability of cognitive processing[2]. At the same time, the introduction of behavioral participation shifted users from passive recipients of information to active participants in the experiential process, enabling them to continuously adjust and reinforce their cognitive outcomes through interactive feedback[3]. Therefore, multimodal interaction not only constitutes an important design principle for digital museum guidance interfaces, but also provides crucial theoretical support for the transition from information display to experience construction.

### *2.2 Sensory Compensation Mechanisms under Visual Dominance*

In human-computer interaction, visual typically serves as the primary channel for information acquisition, undertaking spatial orientation, route navigation and the construction of information hierarchy. However, in traditional digital museum guidance, effective information delivery was often highly dependent on contextual description. Besides, excessive reliance on visual presentation easily leaded to divided attention and increased cognitive load, thereby affecting accuracy of information comprehension[3]. The sensory compensation mechanism refereed to the introduction of other perceptual channels—such as auditory, tactile, and behavioral feedback—on the basis of visual dominance, so as to assist, extend and reinforce visual information, thereby optimizing information transmission through multimodal coordination[6].

Within digital museum guidance interfaces, the visual modality primarily functions to delineate the spatial framework and informational structure, while also providing clear spatial references for other modalities. The auditory modality complements the semantic dimension of visual information through voice narration and environmental sound effects, thereby enhancing the continuity and contextual depth of content interpretation. The tactile modality strengthens users' sense of confirmation and participation through feedback mechanisms such as vibration.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

For example, in the Figure 1, when the interface focuses on a particular decorative detail of a cultural relic, the visual presents its formal characteristics, auditory further explains its historical significance and cultural symbolism, and tactile enhances interactive responsiveness during clicking, zooming, or switching. Sensory compensation under visual dominance is not a simple accumulation of multiple sensory signals, rather than it takes vision as the core index and relies on the timely intervention of auditory and tactile modalities to form a mechanism of sensory complementarity and cognitive regulation[4][6]. In this way, it effectively alleviates the informational burden on the single visual channel and improves both the efficiency of information transmission.

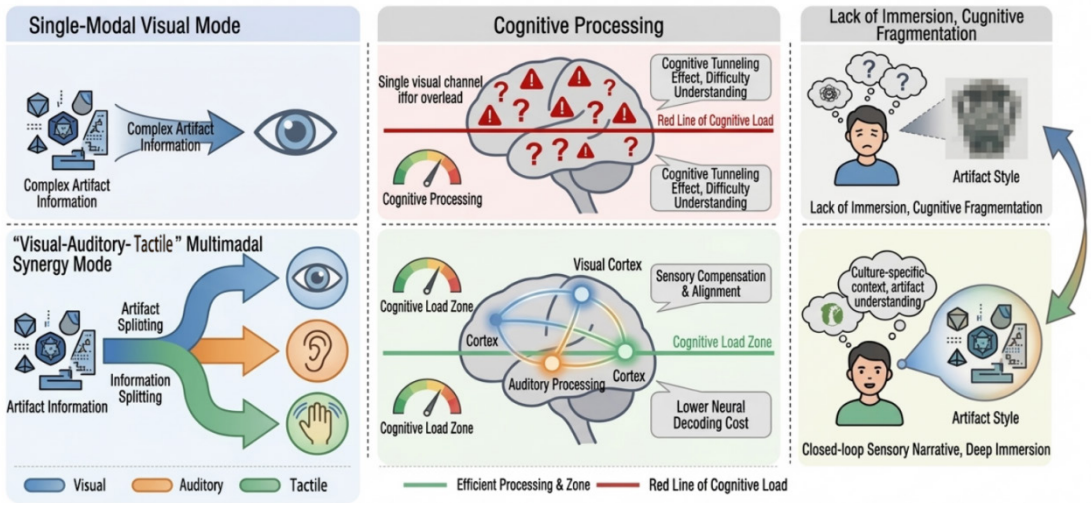



*Figure 1. Multimodal Perception and Cognitive Load Relationship Model Process*

## *2.3 Review of Representative Digital Museum Cases in China and Abroad*

From the perspective of development, both Chinese and international digital museums have gradually introduced the concept of multimodal interaction into the design of guidance interfaces, although differences remain in terms of application priorities and implementation pathways.

Digital museums abroad developed earlier and have formed relatively mature approaches in multimodal interaction, spatial navigation and immersive experience, generally placing strong emphasis on the coordination among visual, auditory and tactile information[5]. For example, the digital guidance systems of the British Museum and the Louvre introduced augmented reality–based spatial positioning and audio interpretation technologies at an early stage. Through the synchronized integration of image display, route navigation and voice narration, these systems enhanced user's continuous perception of exhibition content and their understanding of historical contexts. Such practices implied that international digital museums place greater emphasis on the coordinated integration of the visual and auditory modalities in order to strengthen the narrative quality and immersive character of the guidance process[7].

By contrast, digital museums in China have developed rapidly in recent years and made notable progress in high-precision visual presentation, virtual exhibition and mobile guidance[5]. Represented by projects such as the Digital Duobaoge in Palace Museum and the virtual cave program in Dunhuang Academy, Chinese digital guidance interface systems have reached a relatively high level in the restoration of cultural relic details, the

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

construction of immersive scenes and the experience of interactive browsing[8][9]. Some have also begun to introduce tactile feedback into the guidance process, thereby making preliminary explorations into the implementation of visual–tactile coordination.

However, from an overall perspective, most current digital museum guidance interfaces still exhibit a clear dominance of the visual modality, limited auditory support and relatively weak tactile feedback[4][6]. It indicates that although digital museum guidance practices have been indeed evolving from single-mode display toward multisensory collaboration, further development is still needed in the integrity of multimodal structures and their theoretical grounding. So, these provide a practical basis for the subsequent proposal of the visual–auditory–tactile collaborative model.

# 3. Triangular Structural Model of the Visual–Auditory–Tactile Collaborative Strategy

## *3.1 Theoretical Derivation of the Model*

Within a multimodal information expression system, the collaborative relationships among different perceptual channels directly affect the efficiency of information transmission and the quality of user experience. Traditional digital interface design was often centered on the visual modality, while auditory and tactile channels were usually treated as supplementary components[11]. Single-channel dominance with passive supplementation tends to concentrate informational burden on vision, which may lead to perceptual imbalance, monotonous interaction and insufficient immersion. In response to the limitation, the study proposes a visual-auditory-tactile triangular structural model as a conceptually derived framework for organizing multimodal guidance interfaces in digital museums. The model is grounded in embodied cognition, multisensory learning, and cross-model correspondence research, and aims to explain how different sensory channels can be coordinated to support integrated sensory narration.

The triangular form is used as a structural metaphor to represent the interdependence, complementarity and closed-loop coordination among visual, auditory and tactile. In the Figure 2, although the visual modality often plays a leading role in interface guidance and information indexing, the auditory and tactile modalities are not merely secondary additions, rather, they function as essential channels for semantic extension, behavioral confirmation and emotional enhancement. In this sense, the three modalities form a relatively balanced and mutually supportive experiential structure at the level of experiential organization. Specifically, visual is responsible for informational framing, content presentation and interface guidance, auditory supports semantic supplementation, rhythm organization and contextual atmosphere, and tactile reinforces action confirmation, behavioral feedback and participatory engagement. Through coordinated operation, these modalities jointly construct a complete chain of informational experience and compensate for the expressive and cognitive limitations of any single channel[3].

The theoretical value of this model is summarized in three aspects. First, it emphasizes the relative integrity of multimodal structure: the weakening or absence of any one modality might reduce the completeness of the experiential system and affect the generation of immersion. Second, it highlights the mutual support among

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

modalities, since different perceptual channels are not in isolation, but rather establish bidirectional feedback relationships during interaction. For example, visual would guide tactile intervention, tactile would verify visual operations, and auditory would regulate the rhythm of narration while enhancing perceptual continuity. Third, it stresses the central orientation of multimodal collaboration. The geometric center of the triangle does not correspond to any single modality, but rather symbolizes the integrated sensory narrative generated through the coordinated action of the three channels[4][6]. Therefore, the proposed triangular model should be understood not as an experimentally validated measurement model, but as a theoretically informed conceptual framework for explaining and organizing the structural logic of multimodal guidance interface design.

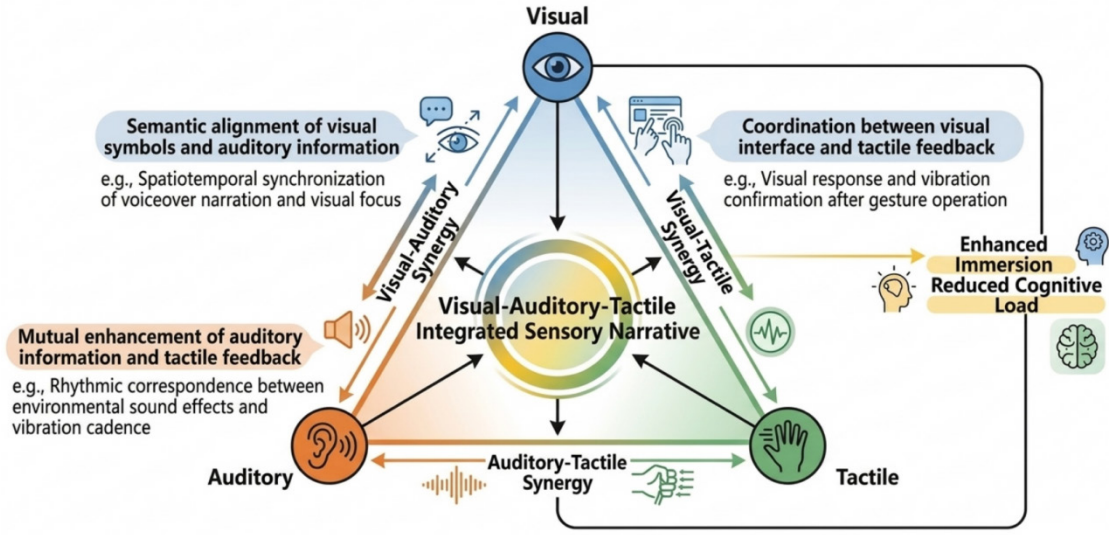



*Figure 2. Schematic Diagram of the Triangular Structure of Multimodal Synergy Strategy*

## *3.2 Logical Mechanisms of Bimodal Collaboration*

Within the triangular structural model, a collaborative relationship can be formed between any two modalities. It forms of bimodal collaboration constitutes the basic unit for achieving multimodal integration[3]. By analyzing the mechanisms of collaboration among different modalities, the internal logic of multimodal information transmission can be further clarified.

### *3.2.1 Visual–Auditory Semantic Alignment*

Visual–auditory collaboration was the most common multimodal combination in digital museum guidance, achieving semantic consistency and temporal synchronization between visual symbols and auditory information[6]. In interface design, visual content—such as images and textual descriptions—and auditory narration needs to maintain correspondence in both spatial position and temporal rhythm, avoiding the cognitive burden caused by informational misalignment[3].

More specifically, visual information should provide the clear structural framework and key cues, while auditory information extends the semantics of the visual content and supplements its contextual meaning. For example, viewing an image of a cultural relic, the simultaneous triggering of relevant auditory narration could enrich the informational dimension without increasing visual burden.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

### *3.2.2 Visual–Tactile Behavioral Mapping*

Visual–tactile collaboration was mainly reflected in the mapping relationship between interactive behaviors of user and interface feedback. In digital guidance interfaces, users interact with the system through actions such as clicking, swiping or route selection, and the interface needs to confirm these behaviors through immediate visual and tactile feedback[3].

Visual feedback is typically manifested through changes in interface elements, such as button highlighting, page transitions or animated effects, while tactile feedback enhances the physical perception of operation through means such as vibration. The collaboration between these two forms of feedback establishes a clear causal relationship between action and response, enabling users to quickly understand the logic of system reactions and thereby improving the controllability and reliability of interaction[6]. In museum guidance systems, visual–tactile collaboration not only enhances interface usability, but also strengthens sense of participation for user, making the interaction process more intuitive and natural.

### *3.2.3 Auditory–Tactile Emotional Resonance: The Synchronous Amplification Effect of Audio Rhythm and Tactile Rhythm*

Compared with the previous two collaborative relationships, auditory–tactile collaboration places greater emphasis on the construction of emotional experience. In multimedia environments, the synchronized relationship between audio rhythm and tactile feedback produce a perceptual effect of resonance at the same frequency, thereby enhancing emotional experience for user[6].

For example, in specific scenarios, subtle tactile feedback coordinated with rhythmic changes in background sound effects would strengthen atmospheric construction, allowing users to obtain more immersive experiences through the combined action of hearing and touch. This multisensory collaboration not only helps increase the affective impact of information, but also improves user's memory of the content[4]. Therefore, auditory–tactile collaboration plays critical role in emotional regulation within digital museum guidance interfaces and constitutes a key component in the realization of immersive experience.

## *3.3 The Closed-Loop Effect of the Trinity*

On the basis of bimodal collaboration, visual, auditory and tactile could form a trinitarian closed-loop structure, achieving an integrated optimization of both information transmission and experience construction.

The closed loop is not a simple superimposition of modalities, rather, it is a continuous experiential process in which users engage at the levels of perception, cognition and action through the coordinated operation of multiple channels[2][3].

First, at the perceptual level, the synchronized input of multimodal information through different sensory channels helps distribute cognitive load and improve the efficiency of information reception. Second, at the cognitive level, the collaborative relationship among visual, auditory and tactile enhances the semantic associations of information, making it easier for users to conduct an integrated cognitive framework. Finally, at the behavioral level, the combination of interactive feedback and contextual experience allows user to receive immediate responses during operation, forming a sustained state of participatory engagement.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

This process was understood as the cyclical mechanism of perception–understanding–feedback. The closed-loop mechanism not only improved the efficiency of information transmission, but also enhanced user's immersion and sense of participation[1][4].

In summary, through bimodal collaboration and trimodal integration, the visual–auditory–tactile triangular structural model realizes a shift from information presentation to experience construction, providing a systematic theoretical framework for the multimodal design of digital museum guidance interfaces.

# 4. Design Principles of Visual Expression Based on Multimodal Collaboration

On the basis of the visual–auditory–tactile triangular structural model, multimodal collaboration is manifested not only in the integration of perceptual channels, but also needs to be translated into implementable design principles. Through the structured analysis of the relationships among different modalities, multimodal visual expression is summarized into three core principles: collaboration, consistency and guidance. These principles enable the coordinated optimization of information-expression efficiency and user experience.

## *4.1 Principle of Collaboration*

The advantage of multimodal expression lies in the expansion of information channels[2]. Therefore, the principle of collaboration emphasizes that, in multimodal design, the functional division of each modality should be clearly defined, avoiding the simple repetition of the same information across multiple channels[3].

In visual–auditory–tactile triangular structural model, the visual modality usually undertakes the tasks of presenting the informational and displaying core content, while the auditory modality supplements the semantic dimension of visual information, and the tactile modality primarily serves to strengthen interactive feedback and action confirmation. By assigning differentiated roles to different modalities, each channel can form a complementary relationship in the process of information transmission, rather than creating redundant repetition.

## *4.2 Principle of Consistency*

The principle of consistency emphasizes the unity of expressive logic and symbolic systems across different modalities, ensuring the stable cognitive expectations within a multichannel information environment[6]. In interface design, visual elements, auditory feedback and tactile responses should be constructed around the unified design language, avoiding interpretive deviations caused by inconsistency in expression.

First, at the visual level, clear symbolic system and interface specification should be established, including color, icons, layout and hierarchical structure. Second, at the auditory level, the tone, rhythm and style of voice narration should remain consistent with the overall style of the visual interface, enhancing the integrity of information expression. Third, at the tactile level, the feedback patterns corresponding to different operational behaviors should exhibit a stable regularity, enabling users to quickly determine system status through tactile signals.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

### *4.3 Principle of Guidance*

In the process of multimodal collaboration, visual usually plays as relatively dominant role. Therefore, changes in the hierarchy of visual information could guide the timing of auditory and tactile intervention, enabling the orderly unfolding of information transmission[3]. The principle of guidance emphasizes the use of visual design to regulate the distribution of attention and trigger auxiliary information from other modalities at appropriate moments.

More specifically, visual contrast, hierarchical emphasis, and dynamic prompts are employed to direct users toward key content. When user's attention is focused on a particular information node, auditory narration is synchronously activated to supplement semantic information, after users complete an operational behavior, tactile feedback is used for confirmation, thereby forming a clear interactive closed loop.

In addition, in complex information environment, strong simultaneous stimulation from multiple modalities should be avoided. Instead, different modalities should be introduced in stages according to user's behavior and contextual states. For example, visual information is dominated during the initial browsing stage, auditory explanation is reinforced during the deeper understanding stage, and tactile feedback is strengthened during interactive operation. Through such visually guided multimodal orchestration, the rhythm and layering of information transmission is effectively enhanced[6].

## 5. Analysis of Multimodal Guidance Strategies in Typical Scenarios

### *5.1 Macro-Level Scenario*

In the virtual exhibition hall roaming scenario of digital museums, users are typically in a state of continuous movement and spatial exploration. As a result, the essential task of guidance design was no longer limited to the presentation of isolated information points, but rather constructing the clear, continuous and directional mechanism of spatial cognition[3].

Compared with static browsing situations, in macro-level scenarios, users rely more heavily on the overall environmental structure, path cues and orientation judgment to complete the process of information acquisition. The role of multimodal collaboration in this scenario is mainly reflected in the organization and reinforcement of spatial perception, for instance, the visual modality establishes the spatial framework, the auditory modality supplements directional information, and the tactile modality enhances behavioral confirmation and physical sensation during movement[5]. Through the coordinated action of these three modalities, continuous movement and spatial exploration could be transformed into a perceptual process characterized by continuity and immersion.

In specific design practice, the visual modality primarily undertakes the functions of spatial construction and route guidance[4]. Through three-dimensional modeling, the organization of perspective relations, the use of wide-angle views and variations in light-and-shadow hierarchy, it helps users quickly form an overall understanding of the exhibition hall structure, while also highlighting exhibition entrances, key exhibits, and navigation nodes through increased brightness, the appearance of interface markers and similar means. The auditory modality assumes the functions of spatial compensation and directional prompting. Through ambient

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

sound effects, voice guidance and spatial audio technologies, it dynamically adjusts the directionality, sense of distance and reverberation depth of sounds in different areas, strengthening user's judgment of spatial zones and movement direction. The tactile modality intervenes through low-frequency, rhythmic feedback, simulating subtle differences in floor materials and movement resistance during path switching, node arrival or the roaming process itself, endowing virtual roaming with a certain sense of physical support and behavioral realism.

It could thus be seen in Figure 3, in macro-level scenarios, the three modalities of visual, auditory and tactile do not function independently of one another, rather, it forms a continuous collaboration among spatial framework, directional perception and movement feedback, improving the fluency, immersion and environmental cognitive efficiency of virtual exhibition hall guidance.

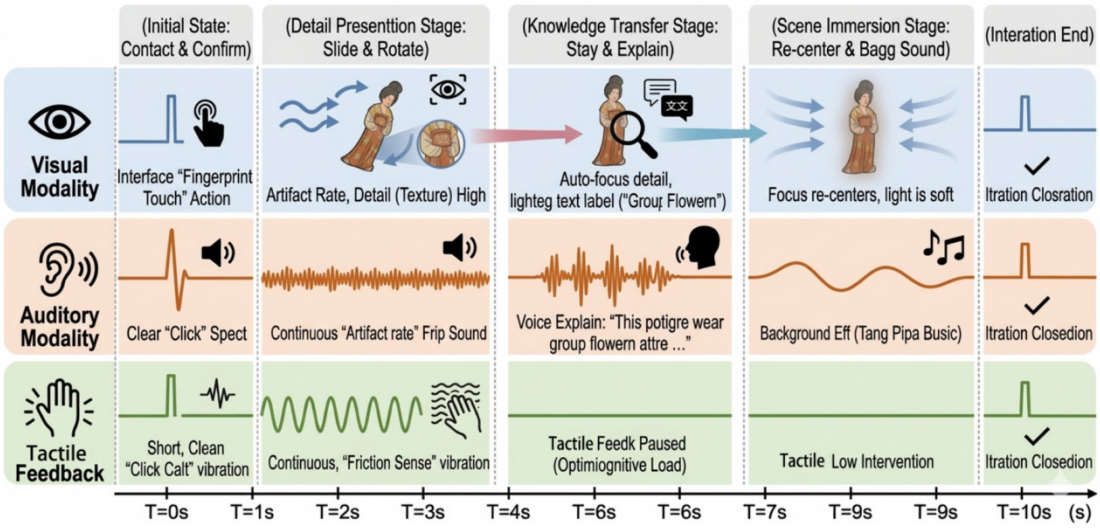



*Figure 3. Three-Modal Interaction Flow Diagram for Digital Museum Artifact Guidance*

## *5.2 Micro-Level Scenario*

Compared with macro-level roaming scenarios, micro-level scenario places greater emphasis on the hierarchical organization of information and the fine-grained coordination of multimodal expression[2]. In the display of specific cultural relics, user's attention shifts from macro-level spatial browsing to close observation, which means design focus of the guidance interface correspondingly should move from route guidance to detail presentation and knowledge decoding. Its essential objective is to help users achieve the in-depth understanding of the details, structural features and cultural significance of cultural relics within a limited interface environment.

On this basis, the role of multimodal collaboration in this scenario was mainly manifested in the following way, the visual modality provides clear detail presentation, the auditory modality supplements the corresponding semantic interpretation, and the tactile modality reinforces interactive behavior and perceptual confirmation, forming a cognitive closed loop of observation–understanding–feedback[3].

To be specific, the visual modality remains dominant by presenting the ornamentation, material, and morphological features of cultural relics through high-resolution images, magnification and structured annotation,

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

while hierarchical information design separates basic explanations from in-depth interpretation. The auditory modality provides semantic supplementation and cognitive guidance by synchronizing segmented voice explanations with user's visual focus, enhancing the comprehensibility of detailed information. Meanwhile, the tactile modality reinforces interaction through immediate vibration feedback and may further simulate material differences through varied intensities or frequencies, offering perceptual support for visual details.

Overall, multimodal collaboration in micro-level scenarios depends not on increasing the number of modalities, but on precise and rhythmically coordinated integration around detail cognition, thereby improving both interpretive depth and interactive quality.

### *5.3 Narrative Scenario*

In historical scenario reconstruction, the core objective of the guidance interface switch from information transmission to emotional experience and narrative construction. Compared with macro-level and micro-level scenarios, its multimodal collaboration is no longer limited to the accuracy of content explanation, but rather lies in enhancing user's immersion, sense of involvement, and emotional resonance through the organization of sensory channels[4][6].

The core of multimodal guidance in historical narrative scenario lies in constructing a closed-loop sensory narrative mechanism, enabling users to transition from content understanding to situational experience through multiple perceptual inputs. The visual modality builds historical scenes through animation, reconstruction and dynamic rendering, providing continuous cues of progression and spatial immersion, the auditory modality enhances narrative atmosphere and emotional expression through music, ambient sound and voiceovers, while the tactile modality extends from operational feedback to emotional support by delivering synchronized vibration aligned with visual and auditory changes. When these modalities achieve consistency in rhythm and emotional expression, forming an integrated sensory synergy that not only facilitates information delivery but also activates users' emotional engagement, deepening the immersive experience of digital museum guidance.

## 6 Challenges and Future Prospects of Multimodal Design in Digital Museums

### *6.1 Technical Constraints and System Optimization*

Although multimodal collaborative design provides a new pathway for optimizing digital museum guidance interfaces, it still faces significant challenges in practical application, particularly at the level of technological implementation. An effective realization of multimodal design depends on the coordinated support of hardware devices and system architecture[5]. However, current tactile feedback technologies remain relatively limited in form, which restricts the role of the tactile modality in emotional expression and informational prompting. At the same time, the synchronized processing and coordinated output of different perceptual channels still present difficulties, while the real-time scheduling of multimodal content also places higher demands on system performance.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

Therefore, future multimodal design practice should further enhance hardware feedback capabilities and optimize the collaborative processing mechanisms of multimodal data, ensuring the stability and fluency of multi-channel information output.

### *6.2 User Experience and Intelligent Development*

While multimodal guidance interfaces enrich the overall user experience, it must also take into account differences in perceptual sensitivity among user groups[10]. Since users vary in their responsiveness to visual, auditory and tactile stimuli, single multimodal combination cannot adequately satisfy diversified demands, especially in terms of accessibility. Therefore, the concept of Universal Design should be incorporated into multimodal design in order to improve system accessibility through multichannel redundancy and substitution mechanisms, while also allowing users to personalize different modalities according to their own preferences[10].

On the other hand, the development of generative artificial intelligence is promoting multimodal guidance systems from static content presentation toward dynamic generation and intelligent interaction[5]. Based on user behavior, such systems can provide personalized content and enhance participation and immersive experience. However, practical application still needs to address issues referred to content accuracy and the norms of cultural representation. Therefore, future research should seek a balance between technological innovation and content regulation, and further explore human–AI collaborative models of multimodal guidance.

## Funding

This paper was the phased achievement of the research project of Jing Chu University of Technology Research Fund Project, An Intelligent Guiding System Study for Museums in the Context of Smart Tourism (YYS202503).

## References

[1]Barsalou L. W. “Grounded Cognition” [J]. Annual Review of Psychology, 2008, 59: 617–645.
[2]Shams L., Seitz A. R. “Benefits of Multisensory Learning” [J]. Trends in Cognitive Sciences, 2008, 12(11): 411–417.
[3]Oviatt S., Coulston R., Lunsford R. “When Do We Interact Multimodally? Cognitive Load and Multimodal Communication Patterns” [C]. Proceedings of the 6th International Conference on Multimodal Interfaces, 2004: 129–136.
[4]Luo D., Doucé L., Nys K. “Multisensory Museum Experience: An Integrative View and Future Research Directions” [J]. Museum Management and Curatorship, 2024.
[5]Li J., Zheng X., Watanabe I., Ochiai Y. “A Systematic Review of Digital Transformation Technologies in Museum Exhibition” [J]. Computers in Human Behavior, 2024.
[6]Spence C. “Crossmodal Correspondences: A Tutorial Review” [J]. Attention, Perception, & Psychophysics, 2011, 73(4): 971–995.
[7]The British Museum. “Audio Guide” [EB/OL]. The British Museum Official Website.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

[8]Dunhuang Academy. “Digital Dunhuang” [EB/OL]. Digital Dunhuang Official Website.
[9]Dunhuang Academy. “Digital Sutra Cave” [EB/OL]. Digital Dunhuang Official Website.
[10]Filová N., Rollová L., Čerešňová Z. “Universal Design Principles Applied in Museums’ Historic Buildings” [J]. Prostor, 2022, 30(1): 92–105.
[11]Wang S. “Museum as a Sensory Space: A Discussion of Communication Effect of Multi-Senses in Taizhou Museum” [J]. Sustainability, 2020, 12(7): 3061.

© 2026 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)