Research on Interactive Mechanism of Multi-modal Environment Design for Immersion Experience

Rui Ye

Yiqinyi Auto Detailing Center, Yuzhong District, Chongqing, 400000, China

Keywords: Multi-modal environment design; Immersion experience; Interaction mechanism; Design principles; Design strategy

Abstract: This article gives an emphasis on the interactivity of multimodal environment design in immersive experience. Multi-modal environment creation is a critical area as it can greatly advance user interaction experience in the age of digital technology. The goal of this work is to investigate how to create a good interaction mechanism in order to increase the level of immersion experienced by users within a multimodal environment. By making full use of the theoretical analysis methods, this article deeply studies the theory foundation of multimodal environment design and immersion experience, and explores the immersions influence of many factors in interaction mechanism on immersive experience like different interaction modes, user behavior need as well as environmental factor. Based on this, this article provides the principles to build interaction mechanism: naturality, consistency, adaptability and introduces how they are realized in detail in design level such as modal fusion, feedback mechanism and situational awareness. Reflection on the principles and strategies from a comprehensive perspective will promote the improvement of interaction quality in multimodal environment design, so that the interaction mechanism is more persevered to user requirements and thus truly provides users with immersive experience.

1. Introduction

With the rapid development of digital technology, various interactive environments have sprung up, and people's requirements for interactive experience are increasing [1]. In this context, the interactive mechanism of multi-modal environment design for immersive experience has become the focus of common concern in academic and practical fields [2]. Multi-modal environment design, as a new field, combines various perceptual modes to create a richer and more natural interactive environment for users [3]. It breaks through the limitation of traditional single-mode interaction, organically combines visual, auditory and tactile modes, and provides users with all-round and multi-level interactive experience [4]. Immersion experience emphasizes the user's full devotion in a specific environment, making it seem to be in a virtual or augmented reality space completely different from the real world, thus gaining a high degree of concentration and emotional resonance [5]. This kind of experience can enhance users' sense of identity and participation in the environment, and also show great application potential in many fields such as virtual reality games, virtual training, digital art exhibition and so on.

From the academic research level, the early research on multi-modal environment design mainly focused on the technical realization level, devoted to how to effectively integrate different perceptual modes into the interactive system [6]. With the deepening of research, scholars gradually realize that simple technical stacking can not naturally create a good immersion experience [7]. Therefore, in recent years, the research has turned to the in-depth exploration of user experience and interaction mechanism, trying to reveal the complex interaction between users and the environment in a multimodal environment, and how this interaction affects the generation and maintenance of immersive experience [8]. However, there are still many shortcomings in this field. On the one hand, the internal relationship between multimodal environment design and immersion experience has not been fully explored, which leads to the lack of systematic theoretical guidance in the design of interaction mechanism. On the other hand, the existing research often focuses on a single mode or

  DOI: 10.25236/ieesasm.2025.017

the combination of a few modes, and there is relatively little research on the synergistic effect between multiple modes and their comprehensive impact on immersion experience.

In practical application, multimodal technology has been popularized in public cultural exhibition spaces such as digital art museum and science and technology museum. However, most of the designs are just a simple stack of technologies, which fails to fully tap the potential of multi-modal interaction in creating a deep immersion experience [9]. For visitors who are mainly teenagers and adults, the existing schemes often ignore their exploration behavior patterns, cognitive load characteristics and emotional resonance needs, which leads to the obvious disconnection between the interaction mechanism and the actual expectations of users. For example, although some exhibitions combine vision and sound, they lack tactile feedback or situational response, which makes it difficult for users to continue to invest. In order to solve this problem, this study focuses on public cultural exhibition scenes, systematically analyzes the internal relationship between multi-modal environment design and immersion experience, and aims to put forward a set of interactive mechanism design strategies with strong operability for ordinary audiences, so as to provide support for improving the experience quality of exhibition space.

2. Theoretical basis and research methods

Multi-modal environment design relies on the cognitive principle of human multi-channel perception integration, and the key is to organically integrate visual, auditory and tactile modes to create a natural, coherent and expressive interactive space [10]. In such an environment, the modes are not simply added, but a synergistic effect is generated by spatio-temporal synchronization and semantic complementation, thus reducing the cognitive burden of users. In a multimodal environment, immersion not only means concentration, but also reflects the user's "Presence" of virtual or enhanced situations, that is, the state in which psychological and physiological responses are closely coupled with environmental stimuli.

From the perspective of cognitive science, human understanding of the external world depends on the integrated processing of cross-modal information. Multi-modal environment design conforms to this mechanism and stimulates users' perception-cognition closed loop through multi-sensory input. For example, vision constructs the framework of spatial structure, hearing guides the direction of attention, and touch enhances the realism of operational feedback. This kind of multi-channel cooperation can effectively activate the joint cortex of the brain and help build a more complete and vivid situational model.

To further explore the mechanism of the above theories in real-life scenarios, this study adopts a research approach that combines theoretical analysis, case studies, and prototype verification. Firstly, through literature review, we will sort out the core concepts and related logic of multimodal interaction and immersive experience. Next, select 5 representative immersive exhibition spaces for analysis, such as teamLab Borderless, the "Digital Scroll" interactive hall at the Palace Museum, the "Interstellar" exhibition area at the Shanghai Planetarium, the "Future Black Box" at the Today Art Museum in Beijing, and the "Deep Sea Exploration" immersive cabin at the Shenzhen Science and Technology Museum. By observing these cases on site, breaking down the interaction process, and recording user behavior, summarize the mapping relationship between multimodal elements and immersive effects. As shown in Figure 1, the immersive experience in a multimodal environment consists of three levels: environmental stimuli, user perception, and experiential outcomes.

Based on this, further research is needed to build a simplified multimodal experience prototype. This prototype is developed based on the Unity 3D engine and integrates the Oculus Quest 3 headset, Leap Motion gesture recognition module, and small haptic feedback device to simulate combinations of different lighting environments, sound field configurations, and tactile intensities. Invite a total of 30 target users to participate in the experience test. Each participant completes three different modal combinations, each lasting 10 minutes, and then fills out the NASA-TLX Cognitive Load Scale and Immersive Experience Questionnaire. The collected subjective data and behavior logs serve as the basis for evaluating the effectiveness of the interaction mechanism.

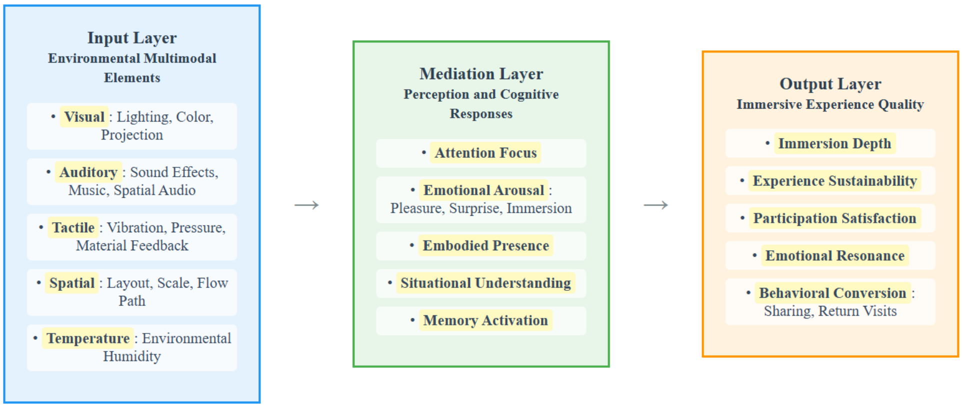

Figure 1 Mechanism model of multimodal immersion experience

3. Empirical analysis of interaction mechanism in multimodal environment design

3.1. Influence of multi-modal collaboration on immersion

In current immersive exhibitions, vision, hearing, and touch are the three most commonly used modalities. If only one of these modes is used alone, it can convey some basic information, but it is difficult to trigger a deep immersion experience. On the contrary, multimodal collaboration will significantly enhance users' sense of presence and emotional investment. Take teamLab's "Crystal Universe" as an example. When the audience enters the exhibition area, the LED lights will dynamically change according to their position, and the background sound effect will gradually transition from low-frequency buzzing to crisp particle sounds. At the same time, the pressure sensor on the ground will trigger slight vibrations. This cross modal consistency is precisely the key to generating immersion.

In order to quantify the effects of different modal combinations, four typical configurations were set up during the prototype testing process, and immersion ratings were collected from 30 users (using a 7-point Likert scale, where 1 represents no immersion and 7 represents extreme immersion). The specific results are shown in Table 1.

Table 1 Average immersion scores under different modality configurations (N = 30)

Configuration	Visual	Audio	Haptic	Mean Immersion Score	SD
A	√	×	×	3.8	0.92
B	√	√	×	5.2	0.76
C	√	√	√ (Low)	5.9	0.68
D	√	√	√ (High)	6.3	0.54

Note: "Low" haptic intensity = 0.5G; "High" = 1.2G. All tasks were identical except for modality setup. Immersion rated on a 7-point Likert scale (1 = not immersive, 7 = highly immersive).

The data shows that only visual mode (group A) has the lowest immersion; After hearing was added (group B), it was significantly improved (+1.4 points, $P < 0.01$). Further introduction of tactile feedback (groups C and D) can continue to enhance the immersion experience, and the higher the tactile intensity, the more stable the score. This proves that modal superposition has marginal gain effect, especially when tactile sensation provides clues to the presence of the body, the immersion tends to be saturated.

3.2. Differentiated expression of user's behavior demand

Multi-modal interaction mechanism should take into account the behavior habits and cognitive preferences of different user groups. During the test, users were divided into two groups according to their age: the youth group (18-25 years old, number n = 16) and the adult group (26-35 years old, number n = 14), and then the differences between them in the choice of interaction mode, operation fault tolerance and feedback expectation were analyzed. The results are summarized in Table 2.

Table 2 Interaction preferences across age groups (% of users selecting each option)

Interaction Aspect	Young Adults (18–25, n=16)	Adults (26–35, n=14)
Prefer gesture interaction	81%	43%
Prefer button/voice support	19%	57%
Tolerate no feedback	38%	14%
Expect immediate multimodal feedback	94%	79%
Average task error rate	12.3%	8.7%

The data show that young users are more inclined to exploratory and non-contact natural interaction, and have higher tolerance for system fault tolerance; Adult users, on the other hand, rely more on clear operation guidance and auxiliary channels and are more sensitive to the lack of feedback. Although the error rate of the youth group is slightly higher, its task completion speed is 23% faster on average, reflecting a stronger willingness to explore.

3.3. The situational moderating effect of environmental factors

Besides modality and users, the physical environment itself is an important variable that cannot be ignored in the interaction mechanism. Such as spatial scale, lighting atmosphere, crowd density, etc., will have an impact on the threshold of users' perception of multimodal stimuli. For example, in the "Interstellar" exhibition area of the Shanghai Planetarium, designers have effectively reduced external interference by reducing the ambient illumination, using omnidirectional sound-absorbing materials and setting up one-way tour streamline, so that the faint starry sky projection and low-frequency cosmic sound effects can be clearly perceived. But if the same content is reproduced in a noisy and bright open square, the immersion effect will be greatly reduced.

Therefore, successful multimodal interaction is not only the integration of technologies, but also the careful design of the environment as a "silent collaborator". Reasonable spatial layout can guide the user's action route and reduce cognitive loss; Soft background light can improve visual contrast; The acoustic environment with low reverberation can ensure the clear transmission of auditory information. These seemingly non-interactive elements are actually the basic support of immersive experience.

4. Construction strategy of interactive mechanism for multi-modal environment design

In order to create a multimodal environment that can effectively enhance the user's immersive experience, it is necessary to follow specific principles and adopt corresponding strategies when designing interaction mechanisms. Naturalness is the physics principle underpinning interactive mechanism design. This principle is a condition for the interaction mode to approach the natural behavior mode of users in real life. With natural play experience, we can save the learning cost of users, help them to focus more on the experiencing content and enhance the sense of being there. The principle of Justice needs to be followed, as well. The interaction logic and feedback mode between different modes in multi-modal environment should be consistent. For example, if a function is triggered by clicking in visual mode, similar operations in tactile or auditory mode should also produce the same result. This consistency can avoid confusion among users, help them form stable interactive cognition, and then better integrate into the multimodal environment. The principle of adaptability emphasizes that the interaction mechanism should be adjusted according to the individual differences and situation changes of users. Different users have different preferences and abilities for interaction, and the interaction mechanism should have certain adaptive ability.

Based on the above principles, the following are some specific interaction mechanism design strategies. Modal fusion strategy aims at organically combining multiple modes and exerting synergistic effect. For example, in the virtual shopping scene, users can not only see the appearance of goods through vision, but also feel the materials through touch, and hear the voice of goods introduction at the same time. Various modes complement each other, presenting goods information in all directions and deepening the user's immersion experience. The feedback mechanism strategy focuses on providing timely and accurate feedback to users. Whether it is operation feedback or environmental state change feedback, it should be clearly perceived by users. Taking the virtual assembly task as an example, when the user successfully completes the assembly of a part, clear feedback is given to the user through visual green prompt, auditory confirmation sound effect and slight vibration in touch, which strengthens the sense of accomplishment and immersion of the operation. Context-aware strategy uses sensors and other technologies to perceive the context information of users, thus providing more suitable interaction. For example, in a multi-modal environment of an intelligent conference room, the system can automatically adjust the brightness of lights, sound volume and the layout of display content according to the information such as the number of participants and the location of personnel, so as to create a comfortable and efficient interactive situation for users and enhance the immersion experience. The interactive mechanism strategy and application of multimodal environment design oriented to immersion experience are shown in Figure 2.

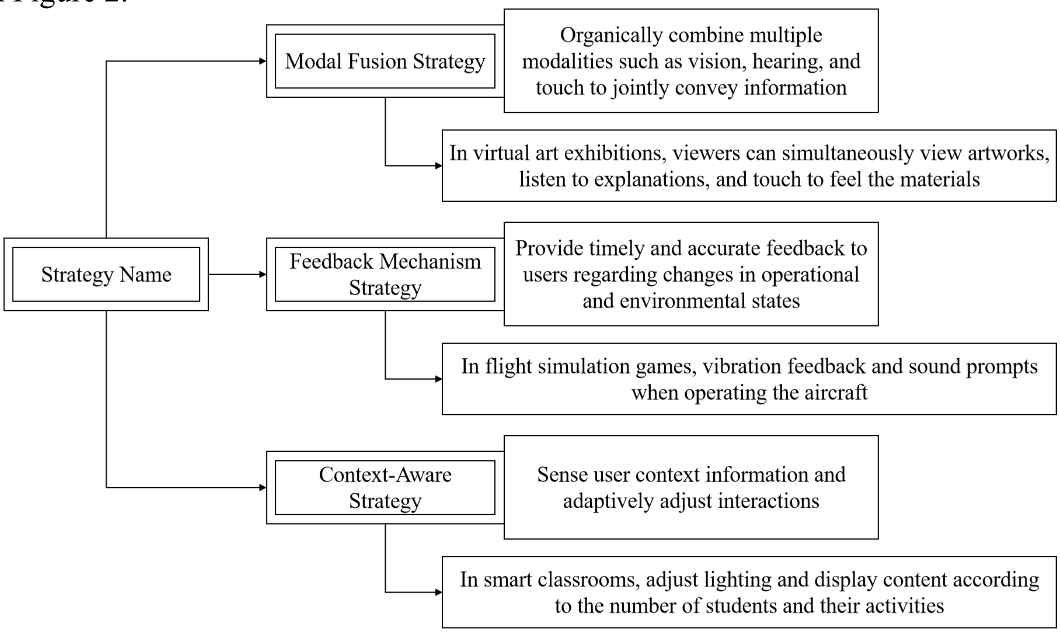

Figure 2 Interactive mechanism strategy for multimodal environment design

By following the principles of naturalness, consistency and adaptability, and using the strategies of modal fusion, feedback mechanism and situational awareness, we can build a more perfect interactive mechanism of multimodal environment design for immersive experience, and bring users a better and deeper immersive experience.

5. Conclusions

This study focuses on public cultural exhibition space, such as digital art museum and science and technology museum, and discusses the immersive experience design for teenagers and adult audiences. Through theoretical combing, typical case analysis and user empirical test based on Unity prototype, this article analyzes the interaction mechanism in multimodal environment and its influence on immersion, and draws the following findings:

The cooperative configuration of vision, hearing and touch has a significant effect on improving the immersion experience. When these three modes are presented simultaneously, the user's immersion score can reach 6.3/7. The higher the intensity of tactile feedback, the more stable the experience.

There are obvious differences in interaction preferences among different user groups. Young users prefer natural interaction methods such as gestures, while adult users rely more on voice or button assistance.

Physical environment is the implicit support of immersion experience. Using spatial strategies such as low illumination, low reverberation and one-way streamline can effectively enhance the user's perception of modal stimuli.

Based on these findings, this article puts forward three interaction principles: naturalness, consistency and adaptability. They are not abstract concepts, but operational guidelines that can be transformed into specific design strategies. Future research results can be directly applied to practical scenes such as digital exhibition optimization, cultural heritage immersion display, and space upgrade of popular science education.

References

[1] Li Xuan, Zhang Ying, Yang Haiyun. Research on Digital Display Design Strategies of Bamboo Culture from the Perspective of Immersive Experience[J]. Journal of Bamboo Research, 2024, 43(02): 71-83.

[2] Zhan Yuhan, Shao Jiang. A Review of Human-Computer Interaction Effectiveness Evaluation Based on Multimodal Interaction[J]. Development & Innovation of Machinery & Electrical Products, 2025, 38(04): 162-165+175.

[3] Luo Xiaofei. A Review of Multimodal Human-Computer Interaction Evaluation[J]. Computer & Network, 2025, 51(03): 258-264. DOI: 10.20149/j.cnki.issn1008-1739.2025.03.009.

[4] Bian Kun, Han Dongnan, Li Siyu, et al. Research Progress in Multimodal Human-Computer Interaction Design[J]. Journal of Machine Design, 2024, 41(11): 199-204. DOI: 10.13841/j.cnki.jxsj.2024.11.025.

[5] Tao Jianhua, Wu Yingcai, Yu Chun, et al. A Survey of Multimodal Human-Computer Interaction[J]. Journal of Image and Graphics, 2022, 27(06): 1956-1987.

[6] Xu Ruyi, Chen Weidong, Zheng Sisi, et al. The Unity of Environment and Body: The Conceptual Construction, Realization Mechanism, and Educational Application of Immersive Experience—Also on the New Field of AI+Immersive Learning[J]. Journal of Distance Education, 2021, 39(01): 28-40. DOI: 10.15881/j.cnki.cn33-1304/g4.2021.01.003.

[7] Mu Chuqiao, Robert Blalack. Self-Discovery and Redesign of the Aesthetic Experience in Immersive Art[J]. Industrial Engineering Design, 2020, 2(06): 67-79. DOI: 10.19798/j.cnki.2096-6946.2020.06.009.

[8] Deng Jiajia, Gong Meijie, Du Zhipeng, et al. Design and Implementation of an Immersive Simulation and Interaction System for Ship Engine Rooms[J]. Chinese Journal of Ship Research, 2023, 18(05): 31-39. DOI: 10.19693/j.issn.1673-3185.02875.

[9] Zhang Xinyu, Guo Lu, Zhou Rongting. Exploring the Business Model of Virtual Reality Publishing—Based on the Value Chain Model of Digital Interactive Services[J]. View on Publishing, 2021, (16): 25-30. DOI: 10.16491/j.cnki.cn45-1216/g2.2021.16.006.

[10] Ge Xin. An Immersive Design Model for New Media Art Experience Exhibitions—Exploring from the Works of teamLab[J]. Design, 2020, 33(15): 42-44. DOI: 10.20055/j.cnki.1003-0069.2020.15.014.