Dean&Francis

ISSN 2959-6157

Research on Somatosensory Audio Tourism and Its Limitations

Yuanzhen Zhang

School of Intermedia Art, China Academy of Art, Hangzhou, China

3200100289@caa.edu.cn

Abstract:

As the demand for gamification continues to rise, the gaming industry is increasingly focusing on motion-based games. The rapid advancement of technologies such as Mixed Reality (MR) and Augmented Reality (AR), along with progress in hand tracking and eye tracking, has introduced a new era of deeply interactive gaming experiences. This study conducts a comparative analysis of existing somatosensory audio games, evaluating the strengths and weaknesses of various technological approaches and proposing innovative solutions to address current limitations, with a particular emphasis on the potential of MR to enhance user experience. The paper explores the potential of integrating technologies like object recognition and gesture recognition to enhance interactivity and convenience in somatosensory audio games. It discusses how such integrations could not only promote game development and innovation but also break down the barriers between the real and virtual worlds. Future research directions are outlined to focus on leveraging these advanced technologies to create more immersive gaming experiences that stimulate and convey music, indicating the significant potential of gamification in the evolution of music and rhythm games.

Keywords: Gamification; somatosensory audio games; mixed reality; object recognition

1. Introduction

Since the inception of Tennis for Two, people have never stopped exploring video games. Whether it's the arcade era when Pac-Man was all the rage [1], the 3D gaming era started by Sony's PlayStation, or the early 21st century when mobile gaming boomed, video games have undergone several changes in how they can be played. In recent years, with the emergence of VR, MR and other technologies, more and more devices that support somatosensory games have appeared in the public's field of vision and have won the favor of players [2]. The development of technologies such as hand tracking, eye tracking, infrared sensing, and motion capture has brought players a rich and colorful gaming experience. *The Infinite Inside* and *Mario Kart Living: Home Circuit* are two of the most successful interactive games on

the market that use MR and AR technologies to achieve a fusion of avatars and real-life scenes. In particular, "The infinite inside" uses gesture recognition technology and eye-tracking technology to realize the interaction between the player's hands and the virtual decryption device in the game without the need for a sensor, and it is presented in real time in the headset.

From ancient tribes to today's high-tech society, music and rhythm play an extremely important role in helping people communicate with each other and express their emotions. However, compared to other categories of somatosensory games, there is still a lot of room for research in somatosensory audio games [3]. Among the existing somatosensory audio games, there are few that break through the two-digit screen while freeing the player's hands, and use MR, object recognition and other technologies to build an immersive experience. What are the differences between the operating mechanisms and technologies of different types of somatosensory audio games? What kind of dilemma is the current somatosensory audio game facing? Is it possible to combine experience in other fields with newly developed technologies to enrich the development of somatosensory audio games? The exploration of these issues is of great significance for the development and user experience of somatosensory audio games.

Based on the comparative analysis of the cases of somatosensory audio games, this study proposes its limitations and possible breakthroughs. This paper aims to classify the existing somatosensory audio games in detail, discuss their advantages and disadvantages from the aspects of game mechanics and related technologies, and point out the enlightenment brought by MR and AR game cases in other fields to somatosensory audio games, and propose to allow users to build an immersive experience with object recognition technology as the core from daily objects.

2. Classification of Somatosensory Audio Games

2.1 2D Screen Class

In the field of games with 2D screens as the carrier, the somatosensory audio games that have been released are roughly divided into two categories, that is, players need to complete the interaction through a device such as a controller, and do not need to grasp the sensing device to score points by body movements. In 2023, SEGA released *Samba de Amigo: Party Central* (Figure 1), a Switch version that takes full advantage of the device's Joy-Con controller and networking capabilities [2] to showcase the physical and "party game" features [3]. The two core themes of "Shaker Dance" and "Samba" are embodied in the gameplay: the Joy-Con (i.e., the shaker in the game) held by the player can accurately recognize the player's body movements, and make judgments on the six strike directions required to determine the score in the game based on the change in the direction and angle of the swing of the hand. In the middle of the song, a series of hilarious and exaggerated action commands will appear, and in order to maintain the combo score, the player must react quickly and use the Joy-Con to strike accordingly.

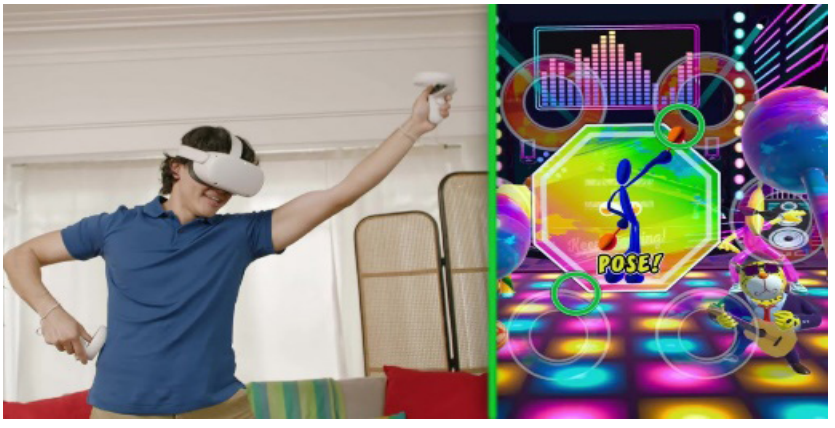

Fig. 1 Samba de Amigo: Party Central gameplay schematic [3]

Samba de Amigo: Party Central is a combination of signature music, moderate cardio, and multi-player "parlor entertainment" that engages the player's entire body to burn fat. In addition, the Switch's networking function allows players to enjoy music in a relaxed and entertaining state in a colorful two/multiplayer mode, promoting communication with others while gaining a sense of accomplishment and belonging. In addition to the social party genre represented by Samba de Amigo Shake Party, there are also games designed for solo players such as *Starri* and *Active Arcade*. Unlike *Samba de Amigo: Party Central*, which relies heavily on gamepads, *Starri* and *Active*

ISSN 2959-6157

Arcade successfully recognize and track the player's body movements through a device with a front-facing camera, such as a smartphone or PC. At the heart of this capability lies AI-powered computer vision technology. In the case of Starri, NEX released its own Motion Developer Kit (MDK) at the same time as the global release of the game. The somatosensory engine developed based on the toolkit uses an efficient scene model and performs far better than the basic human pose estimation in experiments. As a result, Starri is able to more accurately adapt to different body movement speeds and complex dynamics.

NEX's use of computer vision technology has dramatically improved the playability of games, allowing players to start the game anytime, anywhere. This is also an important difference between *Starri* and *Samba de Amigo: Party Central*. Through advanced technology, *Starri* not only provides a more flexible and convenient way to play, but also enhances the user's sense of immersion and engagement, further promoting the development of somatosensory music games.

In addition to the technical differences, in terms of game content, *Samba de Amigo: Party Central* focuses on creating a relaxed atmosphere like an offline music party, while *Starri* and *Active Arcade* focus more on the sense of gain brought by high difficulty and high scores, translating the player's desire to pursue a challenge into actual exercise and testing the results of exercise consumption. For example, Starri uses Nex MDK's proprietary calculation method to provide a more accurate estimate of the calories burned by the player during the course of the song using only computer vision technology. In addition, the ability to reduce background item interference and lock on active players (i.e., allowing people other than the player to appear on the game screen without affecting the player's body determination) is also reflected in the game.

2.2 Immersive VR Experience off Screen - Beat Saber

Like most somatosensory audio games, *Beat Saber* uses the classic "falling" core gameplay, where players need to hit the correct notes to the rhythm while dodging falling obstacles, and cool light effects and virtual lightsabers allow players to fight like Jedi Knight from *Star Wars* (Figure 2). The use of a headset combined with a gamepad is one of the most important features of *Beat* Saber, and what distinguishes it from a 2D somatosensory game like *Samba de Amigo: Party Central*. For example, the HTC Vive, which is supported by Beat Saber, has a sensitive sensor that connects to a display device such as a computer via a Micro-USB cable. In addition, the device is equipped with the Lighthouse location tracking system, which captures the user's free movement within a certain space with the help of spatial positioning technology, and compared to the previous headset, users can move their position and avatar to interact in multiple ways compared to previous headsets.

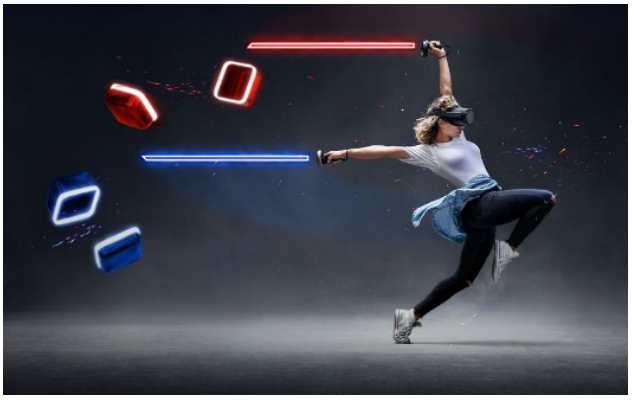

Fig. 2 The official concept image of Beat Saber [4]

The HTC Vive analyzes the position and direction of rotation of the headset and controller in real time through data fusion from two base stations and multiple light-sensitive sensors. This way the devices work together greatly enhances the immersion of *Beat Saber* users. In addition to the device-level benefits, *Beat Saber* is critically acclaimed, the design of the game itself and the creation of the open environment is also essential. First, in terms of note design, unlike the matching of notes and fixed response types in traditional somatosensory sound games, *Beat Saber* adds a kind of inductive notes that can correspond to all directions, which increases the difficulty and fun of the game to a certain extent. The second is the mode of operation, the game dilutes the "touch", and combined with the virtual lightsaber, the player naturally "slashes" to the block representing the musical note [5], which improves the player's sense of playing; The solo mode allows players to customize their library, while a wide variety of mods provide players with a variety of ways to play (such as single-sword gameplay, foot-sensing gameplay, etc.), improving the playability of the game.

2.3 Limitation Analysis

Although the above two types of somatosensory audio games have made certain achievements in their respective fields, there are still some deficiencies in terms of gameplay and equipment. First of all, for two-dimensional screen somatosensory audio games, this kind of game aims to maintain a relatively low threshold for players and show enough convenience, and its user portrait has no excessive restrictions on age, occupation and economic level, etc., and both children and middle-aged and elderly people can participate in it through a mobile

phone under the condition of sports ability. However, this complete separation from reality prevents the use of real-world objects and functions in the game, thus limiting the gameplay of interaction with the joystick/display interface [6], which in turn tends to make the player bored during the game and weaken the durability of the game. In other words, although this type of audio game expands the type of users, it is not good at improving user stickiness. Secondly, many VR somatosensory audio games have a strong dependence on the controller, in addition to meeting the technical needs, the controller is widely used for three reasons: first, the player can control the game through the controller while maintaining a comfortable posture of the body; Second, the operation of the controller is simpler than that of the keyboard, and the player only needs to press a few keys to complete complex game content; Third, from the perspective of the full virtual characteristics of VR, many players need to obtain a connection with reality through the actual controller to meet the sense of security. Although the controller as an extension of the body can bring players a rich experience, on the one hand, the key design of the controller cannot truly restore the function of human hands. On the other hand, the game requires both hands to hold the controller continuously, which can be a challenge for players to use weights for long periods of time, especially for titles like Beat Saber, where the player's hands are waved to determine the score.

3. Improvement of The Use of Somatosensory Audio Game Technology

3.1 Game Development based on MR Technology

The definition of MR technology is generally accepted to be the combination of VR and AR, integrating the real world with virtual elements and digital information. Users can control virtual objects and have them respond to changes in the physical world [7].

The rapid advancement of MR technology has the potential to revitalize the domain of somatosensory audio games by addressing some of the key challenges currently faced. One of the primary issues in audio games is user sensory fatigue, a common limitation that stems from the immersive nature of these games. MR technology, with its ability to blend real-world environments with virtual elements, offers a fresh approach to enhance the gaming experience. Unlike traditional game development that relies solely on virtual constructs, MR introduces a dynamic layer by collecting and processing real-world data, which can be used to create interactive scenarios that are both engaging and novel. This integration of real-world context into the game environment can significantly elevate the playability, offering a more dynamic and varied experience that reduces the onset of sensory fatigue. For instance, consider an MR adaptation of the classic game "Pac-Man," where the familiar maze is superimposed onto a realistic office setting. Players navigate through a maze that is intertwined with their actual workplace, searching for pellets and avoiding ghosts, thus bridging the physical and the virtual. This concept can be extended to combine somatosensory audio games with real-world settings such as escape rooms or playgrounds, creating an intense sensory experience that traditional audio games cannot match.

Moreover, the tangibility and spatial awareness provided by MR can provide a sense of security for players, mitigating the feelings of disorientation sometimes associated with fully immersive VR experiences. This connection to reality can encourage social interaction, as players can engage with each other in a shared space that is both familiar and enhanced by the game.

The current MR game development is mainly based on HoloLens, a headset, and combines hand tracking, hand recognition, eye tracking and other technologies to achieve the purpose of freeing players' hands. It not only satisfies the needs of restoring real hands, but also further enhances the direct interaction between players and virtuals. On the other hand, there are many MR-related cases in the more successful game projects on the market. *The Infinite Inside* and *Mario Kart Live: Home Circuit* are two of the more typical immersive interactive games. The former realizes the integration of virtual images and real things directly through the headset (Figure 3) and enables the player's hand to interact with the virtual decryption device in the game through hand tracking technology. The latter, although popularly referred to as an AR game in a press release, is significantly closer to the definition of MR in terms of interactivity. The game gives players a high degree of freedom to customize the movement of the track and interact with other players' avatars through the Switch (Figure 4).

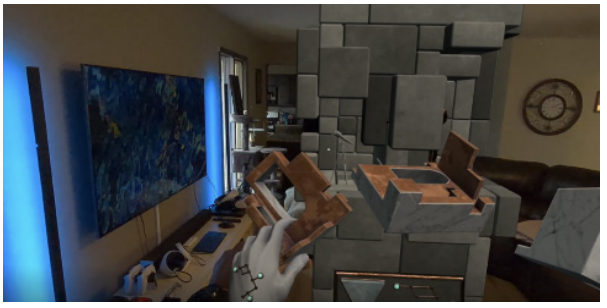

Fig. 3 A combination of virtual and real scenes in The Infinite Inside [7]

ISSN 2959-6157

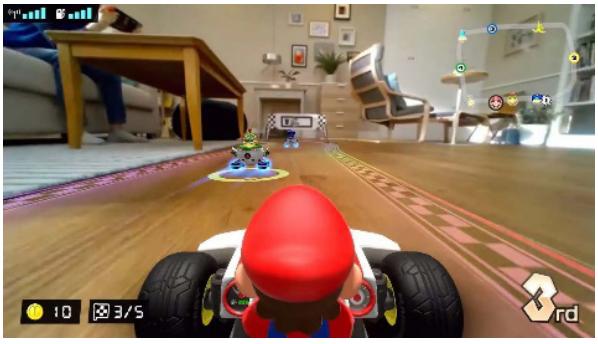

Fig. 4 Mario Kart Live: Home Circuit on the switch [7]

The exploration of MR technology extends beyond recently produced games to experimental endeavors that push the boundaries of the medium. For instance, developers like Volker Paelke and Christian Reimann have crafted an adventure game using the MobEE engine. This game hearkens back to the quintessential 2D adventure gaming experience while integrating innovative MR features. By employing PDA snapshots, it facilitates interactions between NPCs and other elements within a merged reality that overlays the physical and virtual realms. This particular study transcends conventional MR applications that rely on GPS for user location. Instead, it delves into the potential of computer vision technology, proposing the use of mobile device cameras as an interactive medium. Furthermore, it introduces the concept of IOV (Immersive Online Visual) interaction technology to tackle the prevalent issue of latency in MR experiences [8].

The versatility of MR is evident in its applicability to both indoor and outdoor gaming scenarios. It is often amalgamated with other technological advancements to craft distinctive gaming experiences. For somatosensory audio games, this suggests a need to explore innovative convergence points with technology, particularly focusing on gameplay mechanics, hardware requirements, and conceptual frameworks.

In essence, MR technology stands at the forefront of a new era in gaming, offering unprecedented opportunities for immersive and interactive experiences. For the evolving genre of somatosensory audio games, embracing such technology could lead to groundbreaking developments, redefining player engagement and game design.

3.2 Solve the Problem of Convenience and Tactile Sense

MR, with its core characteristics of blending reality with virtual interactions and social elements, holds significant potential for advancing the field of somatosensory audio games. However, there remains a need to address the challenges related to the convenience of operation and the quality of the tactile experience. For instance, games like Mario Kart Live demand additional equipment beyond the base console, such as physical race cars and tracks, which, while offering a degree of freedom, also present a high entry cost and logistical inconvenience for players.

To enhance the immersion of somatosensory audio games, it is essential to consider the sense of touch. The challenge lies in how to maintain the tactile impact of gameplay while simplifying the user experience. Object recognition technology could be pivotal in overcoming these hurdles and setting somatosensory audio games apart from other immersive gaming experiences. Object recognition enables the transformation of real-world objects into in-game elements, allowing players to interact with the virtual world through familiar, tangible items. Presently, object features can be extracted based on color, texture, shape, and spatial relationships, with recognition algorithms having evolved to accurately identify objects using these attributes [9].

Devices like HoloLens, integrated with tools such as Vuforia and MRTK for Unity, use cameras to scan and identify objects in real-time [10,11]. By matching this scanned data with a digital library of models, they can overlay virtual information onto real objects, creating an interactive bridge between the physical and the virtual. This technology could allow players to use common household items to interact with the game, enhancing the tactile experience without the need for additional equipment.

The application of object recognition in somatosensory audio games is an area ripe for exploration and innovation. While these concepts require further experimentation and refinement, the integration of MR and object recognition technologies offers promising avenues for enriching the gaming experience and pushing the boundaries of immersion and interactivity.

4. Conclusion

In conclusion, the domain of somatosensory audio games is ripe for expansion, with advancements in MR and AR technologies leading the way and offering novel research avenues to enhance the immersive experience for players. This paper has meticulously analyzed existing cases of somatosensory audio games, underscoring their merits in liberating players' hands and introducing innovative interaction paradigms. It has also candidly addressed the current limitations regarding technological application, user experience, and game design within these games. Drawing on the experiences of other gaming genres, the paper proposes technological solutions to overcome these limitations. Looking ahead, somatosensory audio games are poised to leverage their unique strengths and integrate

a spectrum of cutting-edge technologies, such as object and gesture recognition. By utilizing everyday objects as interaction medium, these games can amplify players' sense of engagement and deepen their interaction with the gaming environment. This approach is anticipated to not only invigorate the innovation and evolution of somatosensory audio games but also foster a more profound connection between players and the games they play. It is crucial to explore the untapped potential of these games to craft a gaming experience that is both captivating and enriching, propelling the growth of music and rhythm-based games.

The integration of technological innovation with game design is expected to yield a new wave of applications that will significantly contribute to the broader exploration of gamification concepts. It is envisioned that the imminent future will bring forth a plethora of applications that adeptly blend technological prowess with creative game design, marking a significant leap forward in the realm of interactive entertainment.

References

[1] Ye Yuanmao, Luo Hanzhong, Qu Muyan, Wang Zhehao, & Zhang Zhihong. Design and implementation of somatosensory audio game based on rhythm recognition. Computer Knowledge and Technology, 2023, (06): 8-11. .

[2] Cuesta-Gómez, A., Martín-Díaz, P., Sánchez-Herrera Baeza, P., Martínez-Medina, A., Ortiz-Comino, C., & Cano-de-la-Cuerda, R. Nintendo Switch Joy-Cons' Infrared Motion Camera Sensor for Training Manual Dexterity in People with Multiple Sclerosis: A Randomized Controlled Trial. Journal of Clinical Medicine, 2022, 11(12): 3261.

[3] Blaine, T. The convergence of alternate controllers and musical interfaces in interactive entertainment. In Proceedings of the 2005 conference on New interfaces for musical expression, 2005, May: 27-33.

[4] Jiang, H. Design of gesture recognition device for motion sensing game. In AIP Conference Proceedings, 2023, Vol. 3017, No. 1: AIP Publishing.

[5] Dongas, R., & Grace, K. Designing to leverage presence in VR rhythm games. Multimodal Technologies and Interaction, 2023, 7(2): 18.

[6] Buhalis, D., & Karatay, N. Mixed reality (MR) for generation Z in cultural heritage tourism towards metaverse. In Information and communication technologies in tourism 2022: Proceedings of the ENTER 2022 eTourism conference, January 11–14, 2022: 16-27. .

[7] Sargolzaei, P., Rastogi, M., & Zaman, L. Advancing Mixed Reality Game Development: An Evaluation of a Visual Game Analytics Tool in Action-Adventure and FPS Genres. arXiv preprint arXiv:2408.01573, 2024.

[8] Fischer, M., & Wiedner, F. Survey on SR-IOV performance. Network, 2021, 43.

[9] Goel, R., Sharma, A., & Kapoor, R. State-of-the-art object recognition techniques: a comparative study. In Soft Computing: Theories and Applications: Proceedings of SoCTA 2018: 925-932. Singapore: Springer Singapore.

[10] Zhang Maonan, Wang Ke, & Li Peiling. AR game development based on Unity engine. China Information Circles, 2024, (05): 87-89.

[11] Husár, J., Hrehová, S., Knapčíková, L., & Trojanowska, J. Concept of Mixed Reality Application Design for Technical Solutions. In EAI International Conference on Management of Manufacturing Systems: 137-149. 2023, October.