國立臺灣大學電機資訊學院資訊工程學系

碩士論文

Department of Computer Science and Information Engineering

College of Electrical Engineering and Computer Science

National Taiwan University

Master's Thesis

在 VR 體驗中使用基於人工智慧的具象化虛擬助理之研究

Using AI-Based Embodied Virtual Agents in VR Experience

蔡宥杏

You-Shin Tsai

指導教授: 洪一平 博士

Advisor: Yi-Ping Hung, Ph.D.

中華民國 113 年 7 月

July, 2024

doi:10.6342/NTU202402825

Acknowledgements

能順利完成這篇論文，首先感謝洪一平教授從專題生開始一路以來的指導，讓我參與了許多有趣新穎的研究計畫，中途還到德國交換學生一年，在最終論文的研究方向及系統架設也提供了許多建議及指導。在此也要感謝所有 imlab 一起努力過的學長姐以及夥伴。感謝 Bala 花了好幾個半夜陪我理順研究脈絡、感謝喬恩在難熬的日子投餵好大一盒餅乾、感謝郎哥的 KTV 專車、感謝 migu 與冠廷在 314 的日日作陪，感謝靖婷的日更八卦故事、還有一起口試的琪琪、敬媛、呂靜、允玟、世彬以及舒盟，有你們在讓研究及畢業的這條路走起來不再孤單可怕。感謝陪我排練口試的博翔、丞彥，感謝特地來幫我做實驗的 Ian、宥潔還有韓旭。感謝遠在台中的爸媽的無條件支持，還有在口試當天特地來插花的我最愛的姐姐！讓一切在自我懷疑及自我肯定，數著日子順論利利走到今天。

doi:10.6342/NTU202402825

摘要

隨著多媒體技術的進步，人們可以透過虛擬實境（VR）技術遠程探索文化遺址，甚至突破地理限制，參觀已不復存在的歷史場所。過去我們利用敦煌研究院的數據重建敦煌第 61 窟的內部樣貌，讓使用者透過頭戴式裝置觀看這個空間，他們可以選擇敘事性腳本的導覽或自主探索。然而，敘事性導覽可能導致體驗者參與感較低，自主探索則容易讓使用者迷失或錯過重要內容。

本文提出一種融合敘事導覽與自由探索優勢的體驗方法，透過整合虛擬代理人來為使用者提供結構化資訊，並同時保持其自主性。我們使用 Microsoft Azure 和 ChatGPT 創建了對話式虛擬代理人，並在虛擬環境中提供具象化代理人、粒子特效等視覺引導來增強使用者的整體導覽體驗。

基於敦煌第 61 窟虛擬導覽系統的使用者研究顯示，不同虛擬代理外觀（全身、僅頭部、靈魂狀）對使用者體驗有顯著影響。結果指出，全身代理人因其臉部表情和身體動作提供了最高的真實感和存在感，最受參與者喜愛；僅頭部的代理人因形象不完整而不受歡迎，即使其存在感高於靈魂狀代理人。

關鍵字：虛擬實境、虛擬代理人、博物館導覽、對話式互動、大型語言模型

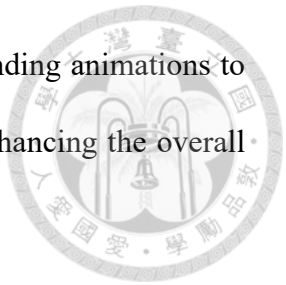
doi:10.6342/NTU202402825

Abstract

With the advancement of multimedia technology, visitors can now explore cultural heritage sites remotely with the support of various technologies. Virtual Reality (VR), in particular, stands out by offering the ability to virtually visit and reconstruct historical sites that are no longer physically present. This breakthrough allows individuals to overcome geographical constraints, enabling broader access to museum experiences. Previously, we harnessed digitized data from the Dunhuang Research Academy to recreate the internal ambiance of the Dunhuang Caves. Through head-mounted displays, users have the option of navigating these virtual spaces either via structured tours or through self-directed exploration. Each approach, however, presents challenges. Structured tours can sometimes lead to a more passive engagement, potentially reducing the user's involvement and the depth of learning. Conversely, the absence of guided tours may leave users feeling lost, potentially missing out on crucial educational content and insights.

This paper introduces an approach to virtual touring that merges the benefits of guided and exploratory experiences by integrating a virtual agent. This method aims to balance the provision of structured information with preserving user autonomy. The proposed assistive guiding system is developed using advanced technologies. Firstly, we utilize Microsoft Azure in conjunction with ChatGPT to create an AI-driven conversational virtual agent. This agent offers users clear guidance and access to extensive knowledge.

doi:10.6342/NTU202402825



Secondly, the system features embodied virtual agents and corresponding animations to provide visual cues and guidance within the virtual environment, enhancing the overall user guiding experience.

The user study was conducted based on the Dunhuang Cave VR experience using the virtual agent guiding system. The experiment aimed to understand the impact of different virtual agent appearances—full-body, head-only, and spirit-like—on user experience during Dunhuang Cave 61 VR tours. We sought to further understand the influence of facial expressions and body movements on users' interactions with the virtual agent. The results show that full-body agents with both facial expressions and body gestures provide the highest realism and presence, and are preferred by most participants. The head-only appearance is the least preferred due to its incomplete human form, although it has a higher presence than the spirit-like agent.

Keywords: Virtual Reality, Virtual Agent, Museum Guiding, Conversational Interaction, Large Language Model

doi:10.6342/NTU202402825

Contents

doi:10.6342/NTU202402825

doi:10.6342/NTU202402825

doi:10.6342/NTU202402825

List of Figures

doi:10.6342/NTU202402825

List of Tables

doi:10.6342/NTU202402825

Chapter 1 Introduction

1.1 Virtual Agents (VAs)

Virtual Agents (VAs), as defined by Vilhjálmsson [35], are characters capable of exhibiting human-like behavior and communicating with users using natural human modalities such as speech and gesture. A classic example of a virtual agent is a personal digital assistant. For example, Amazon's Alexa, Apple's Siri, Google's Assistant, Microsoft's Cortana, and OpenAI's ChatGPT. From the Human-Computer Interaction (HCI) aspect, VAs are crucial in mediating user-system interactions and enhancing the user experience.

In [20], the authors analyzed the development trend of intelligent virtual agents from 2001 to 2015, including 276 user studies in 247 papers. Most of the study focused on discussing the verbal and non-verbal behavior of agents, followed by applications, and at last, physical appearance and identities. They further broadly categorized the applications of virtual agents based on their designed purposes into four, including *virtual assistants*, *training and educational agents*, *social companions*, and *virtual guides*.

Virtual assistants help people finish daily tasks and answer their questions through text or speech, such as voice assistants we mentioned earlier. In marketing scenarios, virtual agents replace real customer service and provide customers with more efficient

doi:10.6342/NTU202402825

and consistent replies.

Virtual agents are also widely used to support *training and educational* purposes. They play simulation or pedagogical roles in these scenarios. For instance, Milcent et al. [19] created virtual patients in the VR simulator for the healthcare professionals' training. Bălan et al. [1] introduced a prototype named eTher for acrophobia therapy in VR, where the virtual agents assume the role of a therapist and continuously encourage the patient throughout the treatment. Petersen et al. [21] conducted an educational VR during the COVID-19 pandemic to introduce viruses through pedagogical agents.

Virtual agents with rich emotional cognition and expression can act as *social companions*. For example, Vardoulakis et al. [34] designed a companion agent to provide long-term and continuous social support to elders. Wanner et al. [38] worked on designing conversational companion agents with linguistic and emotional competence. In recent research, Chaturvedi et al. [8] presented a comprehensive literature review on social companionship in conversational agents and pointed out a few potential research avenues in the future.

Last but not least, virtual agents as *virtual guides* may guide users in the virtual environment through the experience to avoid users for example getting lost. Take a museum guide as an example, virtual guides may enhance users' interest in the artifact through expressive behavior or interactions with users [7] [17]. In addition, in [5], the authors explored user preferences between a virtual guide and free exploration with an embodied conversational agent (ECA) in a virtual museum, concluding that a combination of both approaches may enhance user acceptance.

doi:10.6342/NTU202402825

1.2 Guides for Museums and Cultural Sites

Guidance traditionally involves a trained guide leading visitors through exhibits, offering insights into the history and context of displayed artifacts to enrich visitors' comprehension of cultural heritage. Tours can be broad, encompassing various exhibits, or specialized, focusing on specific themes, periods, or artifact types. With technological advancements, physical guides are no longer the sole means of navigation. Many contemporary guides offer online guidebooks or audio tours to enhance the visitor experience. However, these audio tours typically feature fixed content and lack interaction.

To address these limitations, researchers are exploring ways to deliver a more immersive and interactive tour experience, such as using virtual agents to guide visitors through museums and cultural sites. In the museum context, these guiding agents are defined as a *cyberdocent* [22] or *digital docent* [12], which is the virtual agent that guide users through museums or cultural sites and provide information about the exhibits.

Using multimedia for guiding is a common method, while different displays may be chosen based on different needs—for example, large TV displays [17] [2] [10], immersive projections [26], mobile [12] [28], and head-mounted displays [12] [31] [36]. TV displays and immersive projections usually interact with users through images but lack direct interaction with the physical exhibits. On the other hand, mobiles and head-mounted displays allow users to freely move in a specific space and provide users with corresponding information related to the space.

In addition to different displays, the tour can be categorized into on-site and remote touring. For on-site touring, virtual information is added into the real environment [36]. It

doi:10.6342/NTU202402825

provides users with different aspects of assistance, including augmented information [29], space navigation [11], and entertainment. Remote touring allows visitors to visit the museum or cultural heritage sites remotely, which is essential for the disabled and sites no longer visited. These kinds of guides are popular during the pandemic period due to their virtual visiting feature. In our system, users can remotely enter Dunhuang Cave 61 through a VR headset and freely explore the interior. We aim to address users' disorientation upon entering the virtual space by providing guidance and assistance from a virtual agent.

doi:10.6342/NTU202402825

Chapter 2 Related Work

2.1 Virtual Agents (VAs)

According to the survey on the use of virtual humans in museums and cultural heritage sites [30], there has been a notable shift in research focus from users' avatars to virtual agents over the last five years due to the development of AI. This shift aims to enhance user experiences by fostering social interactions with virtual agents. In support of this, Petersen et al. [21] highlighted that, according to social agency theory, integrating social cues into multimedia lessons can boost learners' social presence, improving cognitive processing and learning outcomes.

The research on the application of virtual agents in guiding is expanding, delving into various aspects such as the agents' visualized appearance [39][24] [21][31], the realism of interaction [17] [5], and natural verbal communication [32][10]. These studies collected subjective and objective measures to understand users' preferences and learning effectiveness. In the following sections, we reviewed the research related to the appearance and interactivity of virtual agents.

doi:10.6342/NTU202402825

2.1.1 The Appearance of Virtual Agents

The visualization of agents has been an important topic, especially in the field of AR and VR [39]. Many studies have discussed this topic from different perspectives, including the degree of visibility [16](whether the users can see the agent or part of the agent), the rendering styles of appearance [41][40] (e.g. abstract, cartoon, or realistic), types of characters [26] (general narrators or story characters), etc. Researchers conducted comparisons from different aspects of virtual agents' appearance. For instance, Petersen et al. [21] conducted a 2 x 2 + 1 between-subject study in the wild and discussed the effect of agents' appearance and behavior realism on learning. Rzayev et al. [23] compared a real word guide with audio-only, audio with a robot-like agent, and audio with a human-like agent in virtual. Sylaiou et al. [31] conducted a study comparing the effect on the users' emotions of the character of storytellers in a virtual museum, including a curator, a guard, or a visitor.

In [26] and [25], Schmidt et al. stated that agent embodiment and thematically related characters induce a higher sense of presence and better user experience in VR exhibition space, which has been verified in their experiments. However, the impact on learning is inconclusive. Some researchers considered positive, some negative, and some suggested that it is related to the type of information, being conceptual or factual [39]. Based on these studies, we believe that virtual agents capable of providing users with a higher sense of presence or social presence can generally enhance users' learning of the tour content. Therefore, in our user study, we will explore how varying degrees of humanoid appearance in virtual agents impact users' sense of presence and social presence in museum-guiding scenarios, particularly under conditions of spatial guidance and conversational interaction.

doi:10.6342/NTU202402825

This type of scenario has been uncommon in previous research, highlighting the importance of our investigation in understanding how virtual agents can improve user experience in such contexts.

As shown in the Figure 2.1, we can separate the agent's appearance into non-embodied and embodied. Agents without an embodied appearance typically interact with users only through voice or text, such as chatbots or personal voice assistants. On the other hand, agents with an embodied appearance can be further classified based on whether they have a humanoid appearance. For example, as discussed by Wang et al. in [37] in the context of AR, a "smart home" device is an agent without a human-like appearance, and a full-body agent has a human-like appearance and further enhances user experience by simulating human behavior, providing an interaction that closely resembles real human interaction.

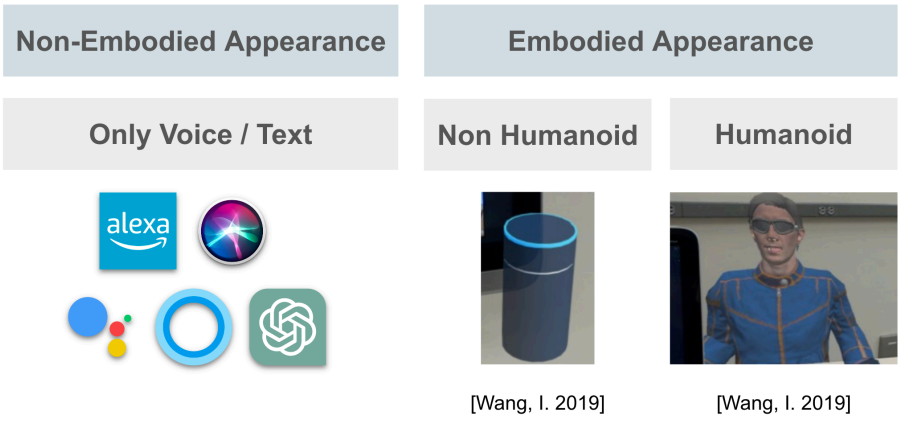


Figure 2.1: Virtual Agent Appearance Classification

2.1.2 The Interactivity of Virtual Agents

In general, virtual guides are one-way interactive, with the docent introducing the tour and the visitor listening. To enhance the overall experience, the interaction between the virtual agent and the visitors is crucial, for instance, providing personal explanations or reactions based on users' responses and behavior [5] [17] or making virtual agents with

doi:10.6342/NTU202402825

the ability to talk naturally to users and reply to their questions [33][10][7]. In summary, as shown in Figure 2.2, interactions between users and virtual agents can be broadly categorized into two types: one-way and two-way interactions. One-way interactions are typically used in pre-recorded guides, for example, [31] and [26] where the virtual agent provides information to users. In contrast, two-way interactions are more interactive, allowing users to ask questions or interact with virtual agent through body gesture, as seen in [10] and [17].

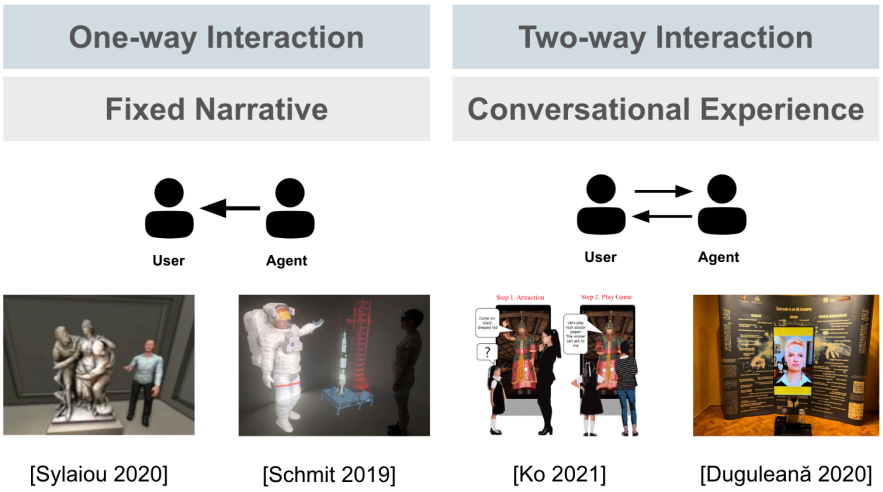


Figure 2.2: Virtual Agent Interactivity

2.1.3 Using Virtual Agents in Guiding Experience

According to the overview of virtual agent studies by Norouzi et. al. in [20], virtual agents have been used in many areas and perform numerous functions, such as accompanying, teaching, guiding, assisting, and so on. As illustrated in Figure 2.3, various types of virtual agents are used for guiding, depending on their appearance and interaction methods. Pre-recorded audio guides typically fall into the bottom-left group, while text-based chatbots or voice assistants, which offer more user interaction, are in the bottom-right group. Virtual agents with embodied forms are often utilized in VR experiences for spa-

doi:10.6342/NTU202402825

tial and plot guidance in the virtual environment, positioning them in the top-left space. For enhanced user interaction, many embodied virtual agents in the top-right group are equipped with verbal or non-verbal interaction capabilities.

In the past, Dunhuang Cave 61 was a fixed narrative VR experience with voice-over, which can be considered a non-embodied agent. In our study, we aim to transform Dunhuang Cave 61 into a conversational VR experience with an embodied agent appearance through the establishment of a virtual agent guiding system, falling into the upper right quadrant of Figure 2.3.

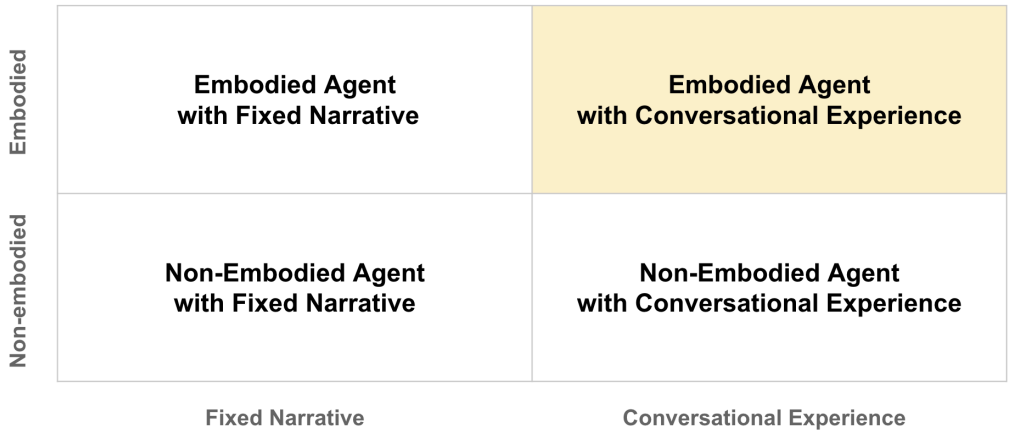


Figure 2.3: Virtual Agents in Guiding Experience

2.2 Using Large Language Model in VR

Artificial Intelligence (AI) has seen extensive use in Extended Reality (XR) in recent years for various purposes [15]. For example, creating XR worlds, understanding users, and supporting interaction in XR. Following last year's breakthrough in generative AI, large language models (LLMs) are now being applied across numerous fields, including VR and AR. For instance, [9] demonstrates the use of LLM to implement a virtual assistant capable of verbal communication, aiding participants in responding to anatomy

doi:10.6342/NTU202402825

questions of varying cognitive complexity. In addition, many VR or AR applications (e.g., games and teaching platforms) use LLM to drive virtual characters to communicate with users. In our study, we apply LLM technology to virtual tour scenarios, aiming to address the common bottlenecks in VR experiences by enabling conversational interactions with embodied virtual agents.

doi:10.6342/NTU202402825

Chapter 3 System Design

3.1 System Overview

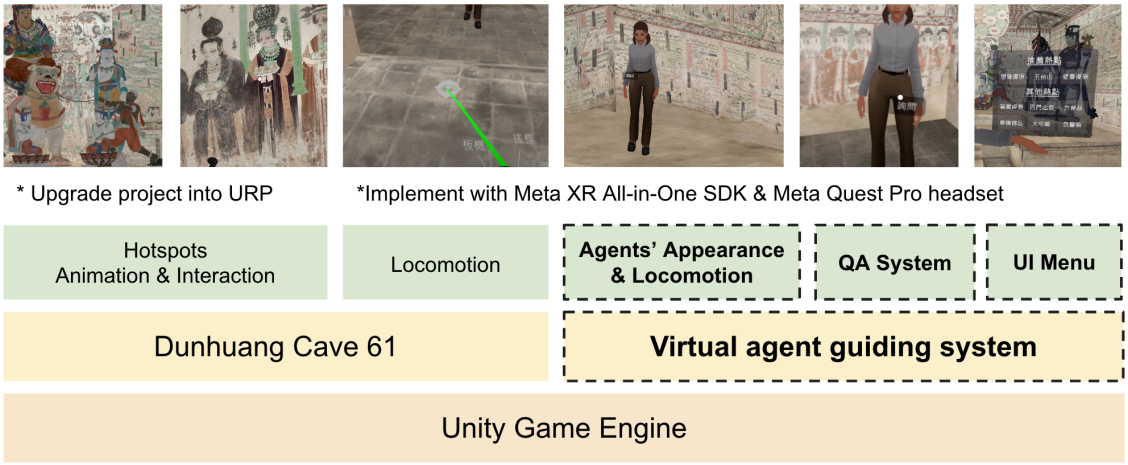


Figure 3.1: System overview

As depicted in Figure 3.1, the whole VR experience is developed in the Unity game engine. The original Dunhuang Cave 61 in VR includes numerous hotspots, such as watchable animation and interactive murals. For better performance and cross-platform compatibility, we upgraded the original project to the Universal Render Pipeline (URP) version. We developed a virtual agent guiding system to guide the user in cave 61, as the blocks in Figure 3.1 are enclosed in dotted lines. According to its function, it can be divided into three parts: the QA system, the visualization and locomotion of the agent, and the UI menu interface.

doi:10.6342/NTU202402825

3.2 QA System

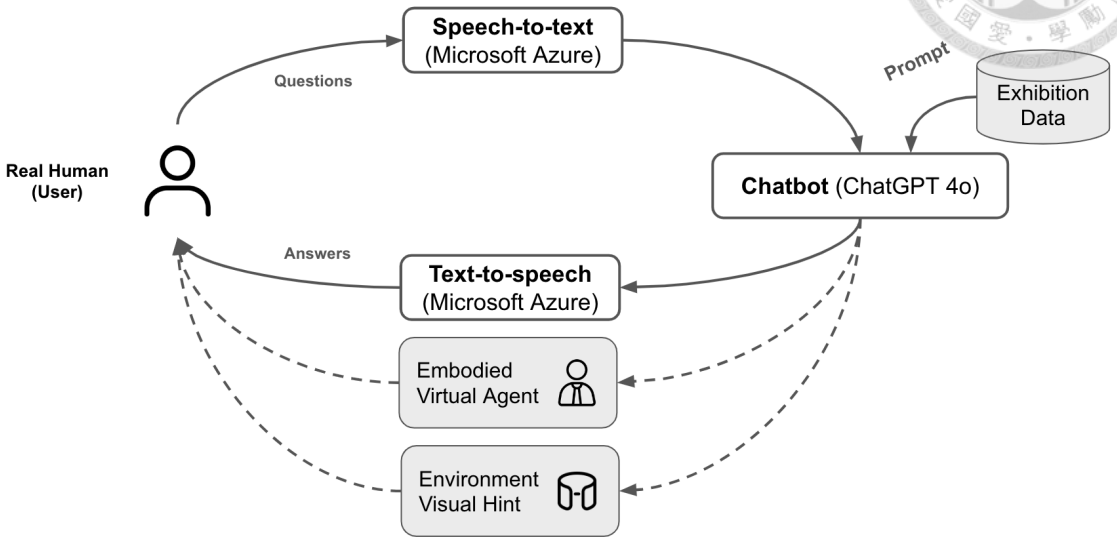


Figure 3.2: Interaction loop of conversational AI in VR environment

To enable the virtual agent to interact with users, we integrated generative artificial intelligence (AI), text-to-speech (TTS), and speech-to-text (STT) technologies. As illustrated in Figure 3.2, users can ask questions via a microphone. These spoken queries are converted into text by Microsoft Azure Speech Service. The resulting text is then processed by the large language model (LLM), OpenAI ChatGPT 4o. Finally, the generated responses are converted back into speech by Microsoft Azure and delivered to the user. The whole conversation was in Mandarin.

We used single-shot recognition during the voice recognition process to ensure clear and fluent user interaction, meaning the system recognizes one utterance at a time. Recognition starts when the user triggers the ask button and stops when a silence is detected or a maximum duration of 15 seconds is reached. By default, the system allows an initial silence of 5 seconds before determining the recognition result as ”no match.” Additionally, the default value for segmentation silence, used to determine the input as ”done,” is

doi:10.6342/NTU202402825

0.5 seconds. This means that if the user pauses for more than 0.5 seconds, the system will terminate the current voice recognition session.

To improve the quality and accuracy of the responses, we used the prompt engineering technique for LLM. We provided the model with existing text context about Dunhuang and Cave 61. This context includes the history of the caves, details about their murals, and information on all hotspots featured in the existing VR experiences of the cave. Additionally, we defined the virtual guide's personality and response style in the prompts, hoping to give the user a professional and friendly image.

3.3 Embodied Virtual Agents

3.3.1 The Appearance and Behavior of Virtual Agents

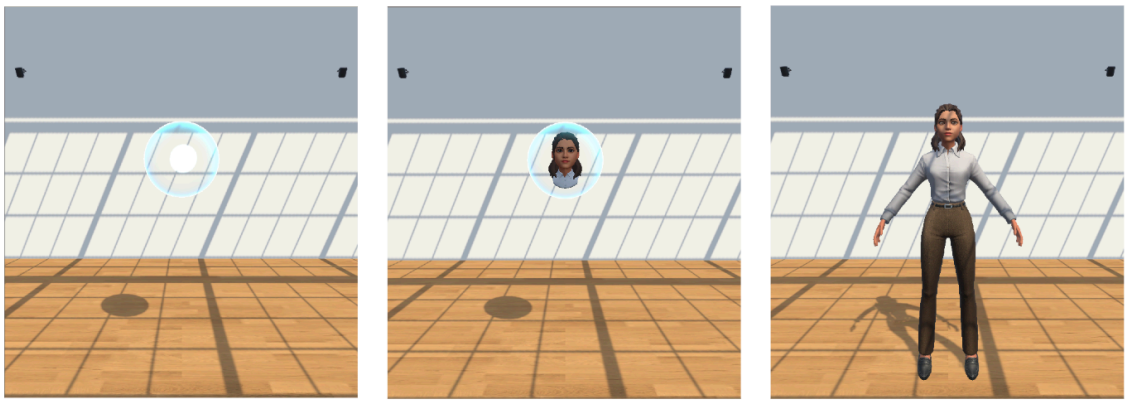


Figure 3.3: Different kinds of virtual agent's appearance, from left to right: Spirit-like, Head-only, and Full-body.

Based on the style of Dunhuang caves and the application of virtual agents, we designed three different appearances, namely, spirit-like, head-only, and full-body (See Figure 3.3). The head-only and full-body appearance is built by Read Player Me, and their appearance is chosen to be a female human wearing a blouse and trousers to express friend-

doi:10.6342/NTU202402825

liness and professionalism. To minimize the unnatural look of the head-only appearance, the head-only model is put in a transparent bubble.

All three kinds of appearances support text-to-speech functions, which is a part of the QA system. In addition, all three kinds of appearances have particle effects when guiding users to specific hotspots. However, only head-only and full-body appearance support facial expressions, such as lip-sync and eye blinking. Additionally, only full-body appearance includes human body gestures and animations, the other two only contain animation of floating bubbles.

3.3.2 The Position and Orientation of Virtual Agents

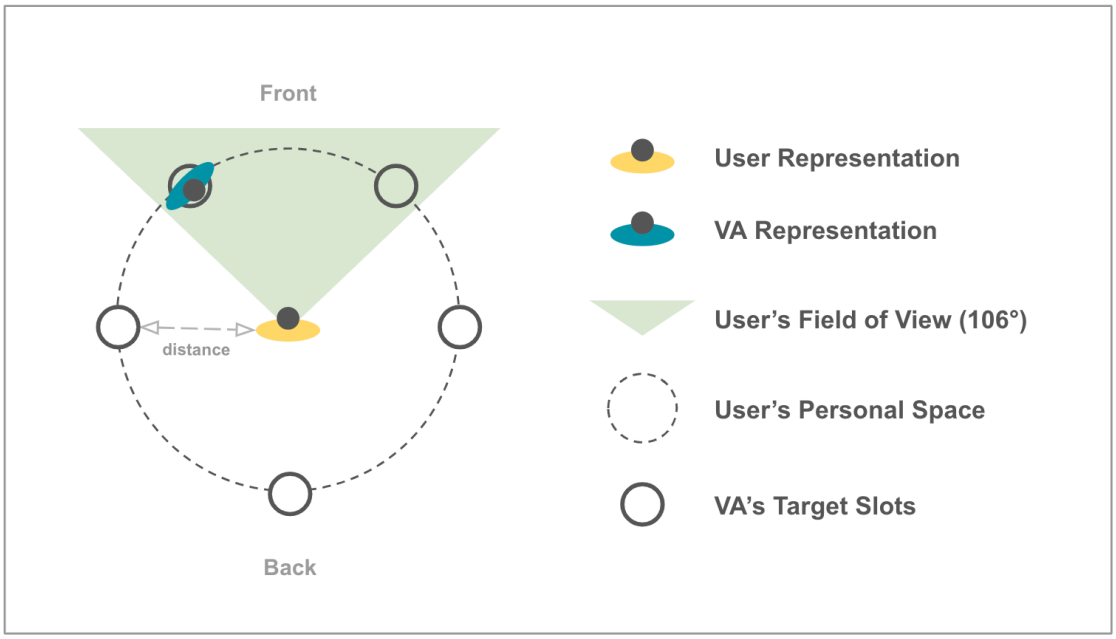


Figure 3.4: Target positions of the virtual agent in relation to the user

The location of the virtual agent follows two rules. First, when the user requests the virtual agent to introduce a specific hotspot, the virtual agent will first teleport to the best viewing point of the hotspot and invite the user to come over. Otherwise, the virtual agent will always accompany the user. Once the users need assistance, the virtual

doi:10.6342/NTU202402825

agent appears near them, giving them chances to ask questions. To provide users with a comfortable and natural social interaction experience, interpersonal distance, defined as personal space (PS), is considered. In [13], Hall divides the interpersonal distance into four categories, ranging from intimate(0-0.45 m), personal(0.45-1.2 m), social(1.2-3.6 m) to public(> 3.6 m). Therefore, we define the virtual agent's target slots to be distributed 1.5 meters away from the user's position (see Figure 3.4), which is slightly larger than the minimum social distance to reduce the feeling of oppression. Five target slots are respectively on the left front, right front, left-hand side, right-hand side, and back of the user. In addition, to ensure that the user can directly see the virtual agent, but the virtual agent does not block the user's main line of sight, the left-front and right-front target slots are located near the peripheral edges of the user' s field of view (FOV) when wearing the head-mounted display. When the user teleports in the virtual environment, the system decides the position in the above order. The next position will be considered when the previous option is invalid, such as being out of space.

3.3.3 Virtual Agent State Diagram

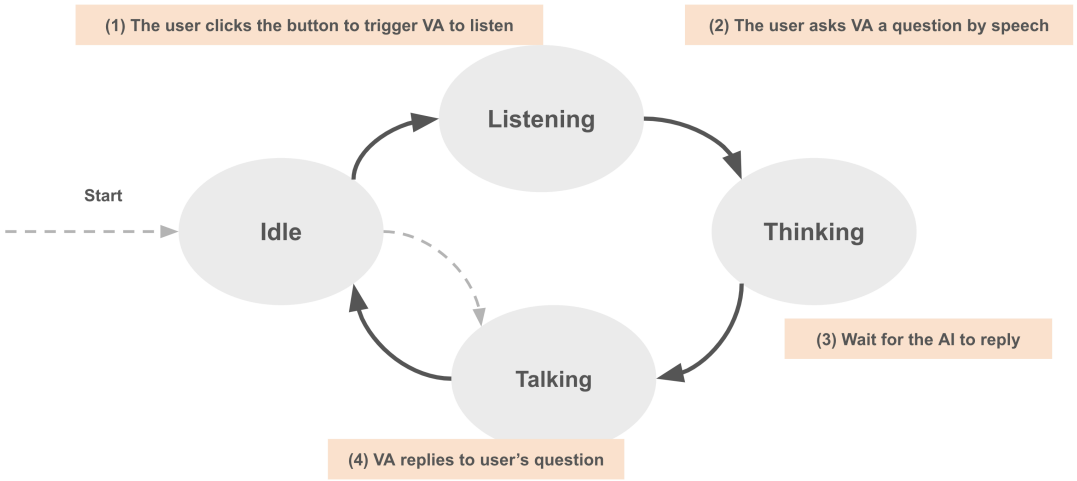


Figure 3.5: Virtual Agent State Diagram

doi:10.6342/NTU202402825

The virtual agent operates in four distinct states: idle, listening, thinking, and talking, as depicted in Figure 3.5. The transition between these states occurs based on the user's action and the virtual agent behaves differently depending on its state.

As illustrated in steps 1 to 4 of Figure 3.5, when the user triggers the virtual agent, the user can ask questions via voice. Then, the user's speech will be converted into text and sent to the language model for processing, which is the thinking state. Finally, in the talking state, the textual responses from the language model are transformed back into speech and delivered to the user. If the user keeps asking questions, the virtual agent will be in the above loop. We use character model animations sponsored by Reallusion Inc. to ensure that the full-body agent has corresponding different body animations in four different states.

3.4 UI Menu and User Manipulation

Figure 3.6: Full-body Virtual Agent with Menu in Dunhuang Cave 61

doi:10.6342/NTU202402825

3.4.1 UI Menu

The user can press the A button on the right-hand controller to trigger the UI Menu to appear or hide. As shown in Figure 3.6 Users can not only ask the virtual agent for the location of a hotspot, such as ”Where is the Wutai Mountain map?”, but also directly select the desired hotspot from the UI menu on the right side of agent. After that, the virtual agent will teleport to the best viewing point of the attraction and then invite the user to join her. Both methods trigger particle effects from the user's handheld controller to the target hotspot, serving as a visual guide for the user.

3.4.2 User locomotion

The physical size of Dunhuang Cave 61 is quite large, and the murals are all over the cave. Constrained by physical space limitations and to avoid users' VR motion sickness, non-continuous teleportation-based movement is chosen [4]. In this VR experience, users can instantaneously teleport to a chosen position with a specific orientation by the joystick on their right-hand controller.

doi:10.6342/NTU202402825

Chapter 4 User Study

To better understand how the visualized appearance of embodied virtual agents affects users' experience, we conduct a with-in-subject experiment with three conditions: virtual agents with spirit-like, head-only, and full-body appearance (Figure 3.3).

4.1 Participants

This study was conducted with 18 participants (10 males and 8 females) ranging from 21 to 29 years old (mean = 24.83, SD = 2.04). Seventeen participants had prior experience with VR, and nine of them had experience in developing VR applications. Two participants had no experience in using voice assistants; the other sixteen participants did, and five of them used them regularly.

4.2 Experimental Scene

Figure 4.1 is a top view of Dunhuang Cave 61, which serves as the experimental scene for this study. The locations numbered 1 to 9 in Figure 4.1 represent nine hotspots within the cave, each featuring some animations or interactions.

doi:10.6342/NTU202402825

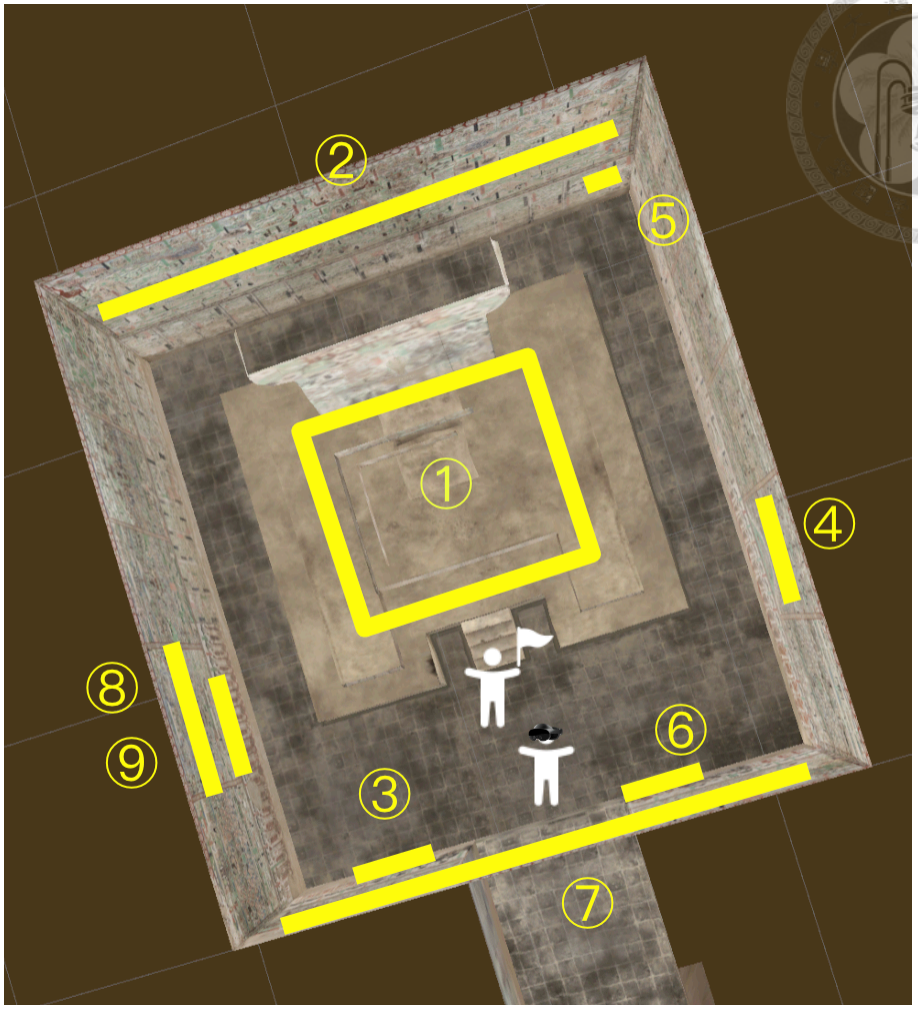


Figure 4.1: The top view of Dunhuang Cave 61

4.3 Measures

The user subjective evaluation of this experiment consisted of three parts: overall user experience, interaction with the virtual agent, and system usability. For overall user experience, we used the igroup presence questionnaire (IPQ) [27] to measure the user's sense of presence. Regarding interaction with virtual agents, six questions are derived from the study by Harms et al. [14] to measure three of the dimensions of social presence, including co-presence, attentional allocation, and perceived message understanding. In addition, the study by Wang et al. [37] to evaluate other aspects of virtual agents are also included, including helpfulness, presence, relatability, trust, distraction, and realism. For

doi:10.6342/NTU202402825

the overall system evaluation, we utilized the System Usability Scale (SUS) [6] for system usability evaluation. Participants were also asked to rank the three appearances of virtual agents in order of preference.

4.4 Procedure

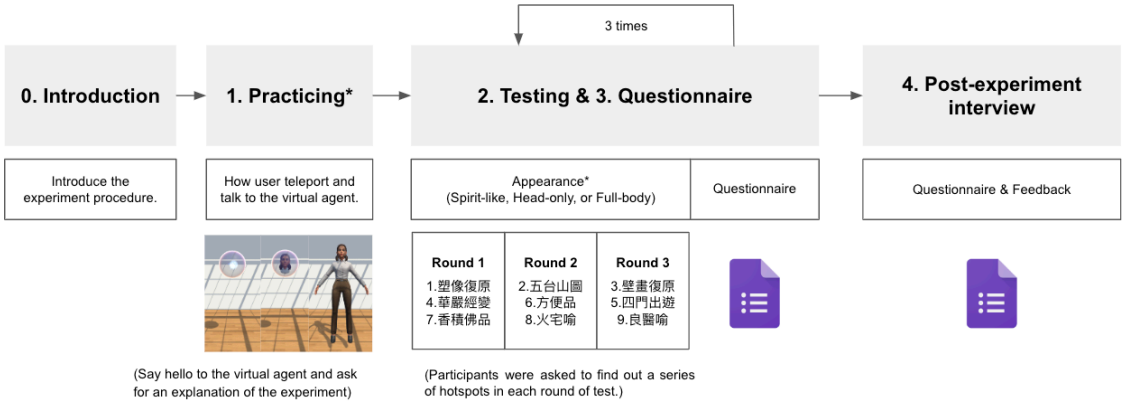


Figure 4.2: Procedure of the study about appearance.

Figure 4.2 presents an overview of the study procedure. The order of the conditions during the testing session and the appearance of the virtual agent used in the practicing session are counterbalanced (The conditions marked with * were counterbalanced). The study begins with the experimenter introducing participants to the whole experimental procedure. This is followed by a practicing session in which participants learn to teleport and interact with a virtual agent in VR with right-hand controller. During the main testing session, participants will experience three different conditions and then fill out separate questionnaires. This was repeated three times, and each time participants were assigned a different task, such as using the virtual agent guiding system to find specific hotspots. Before the study ended, a post-experiment questionnaire and an interview were conducted to collect feedback and suggestions from participants.

doi:10.6342/NTU202402825

4.5 Result

4.5.1 System Usability

To evaluate the performance of the guiding system, SUS was conducted. The results of each question in the SUS questionnaire are shown in Figure 4.3. The even-numbered questions have been reversed. Except for Q4, all the questions get over 5 points on average on a 7-point rating scale. We speculate that the average score for Q4 is lower because not all participants are very familiar with the control of VR controllers. The final SUS score (0 - 100) on average is 78.8 (SD = 9.14).

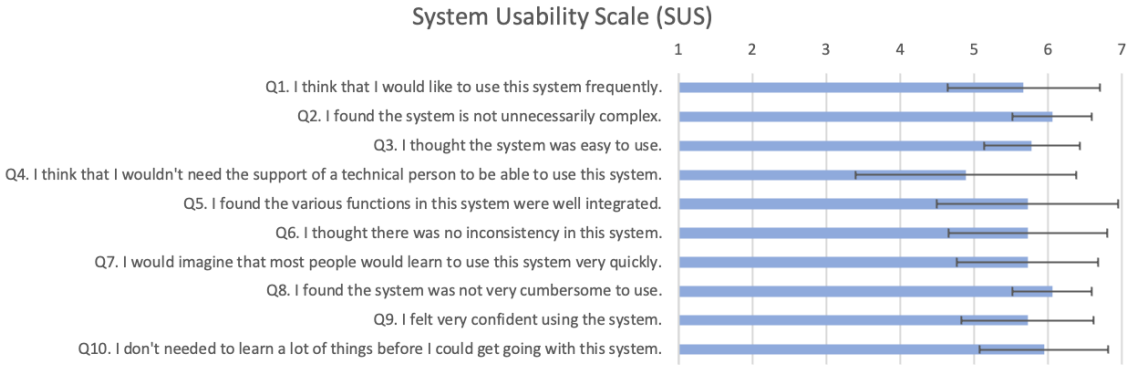


Figure 4.3: The results of system usability scale

4.5.2 Sense of Presence

In [18] conclude the concept of presence in VR as a subjective perception of ”being” in the virtual environment, which is also a main measure of uccess both in applied and academic work on VR. The results of users’ sense of presence are shown in Figure 4.4. The results showed consistent effects and no significant differences between the three looks on all four dimensions of presence (including sense of being there, spatial presence, involvement, and experienced realism).

doi:10.6342/NTU202402825

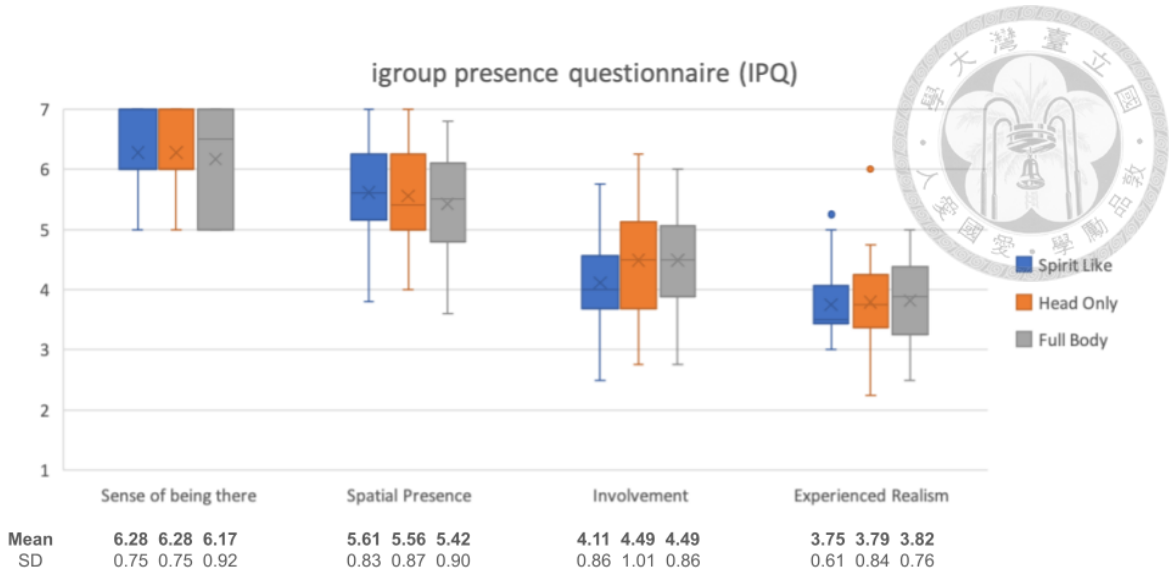


Figure 4.4: The results of IPQ

4.5.3 User Perception of Agent

We ran ANOVA on six scales individually to find differences between three agent appearances. Graphs comparing agents for each dimension are depicted in Figure 4.5. We found that agents' appearances have a significant effect on the presence, trust, distraction, and realism ratings. For the presence and realism rating, the full-body agent had a significantly higher score than both spirit-like and head-only agents. For trust and distraction ratings, full-body agents had significantly higher scores than spirit-like agents.

Social presence, defined as the "sense of being together with another" [3], indicates the users' perception of agents' presence in the virtual environment. To further analyze agents' presence, we extract six questions in three dimensions from the [14]. The results are shown in Figure 4.6. We found only a significant difference on the second question: " The virtual agent's presence was obvious to me.", where the full-body agent had a significantly higher score than the spirit-like and head-only agent.

doi:10.6342/NTU202402825

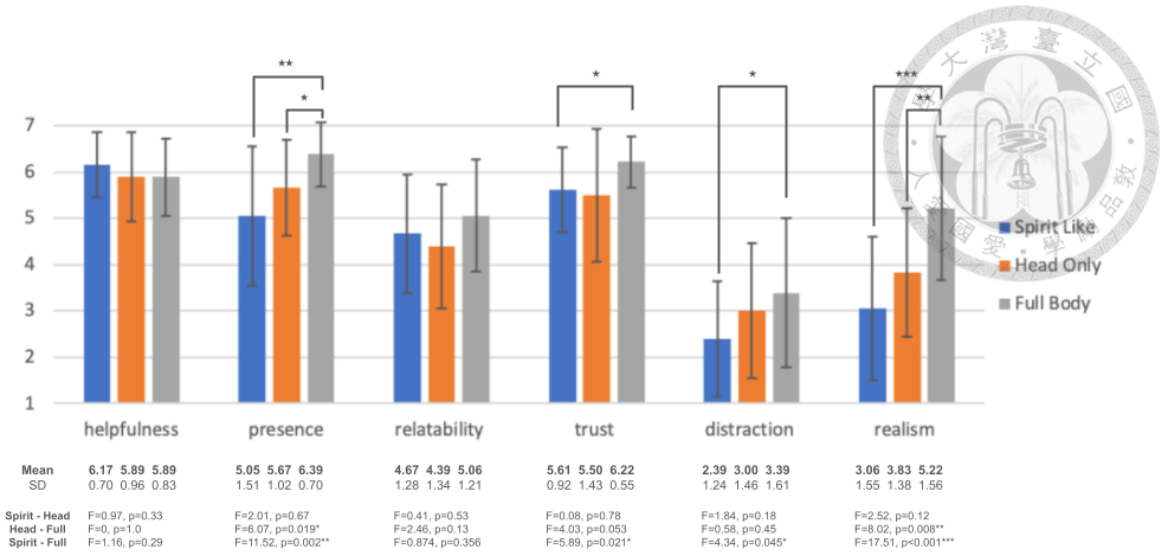


Figure 4.5: The results of users' perception on agent in six dimensions

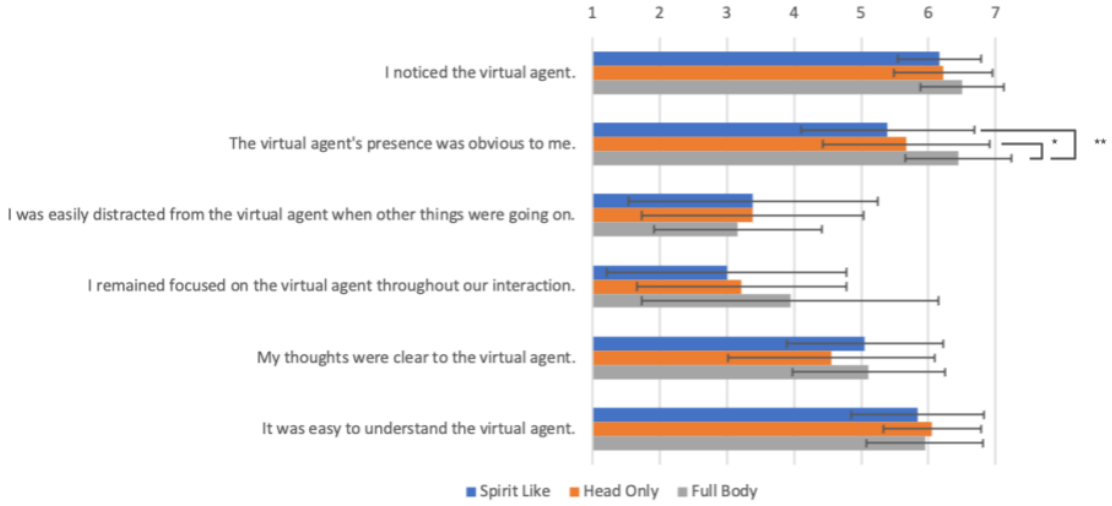


Figure 4.6: The results of users' perception on agent in social presence

4.5.4 User Preference

To understand user preference for the three different agent appearances, we asked users to sort three appearances according to their personal preferences. The result is shown in Figure 4.1. The full-Body agent had the most first-place favorite, with a total of 10 votes. Then the spirit-like agent received 5 votes and finally, the head-only received 3 votes.

doi:10.6342/NTU202402825

	Spirit-Like	Head-Only	Full-Body
Favorite	5	3	10
Second Favorite	9	6	3
Third Favorite	4	9	5

Table 4.1: The results of participants' preference ranking on appearance

4.5.5 User Reviews About Agent

To further understand why each participant prefers one appearance over another, we look at the feedback of the opening question asking them about their sorting reason and the post-experiment interview. We summarize their comment on each appearance into positive and negative perspectives.

4.5.5.1 Spirit-Like Agent

Participants used *"sense of technology"* (P7, P8) to describe their feelings toward this agent, and one participant who favored the spirit-like agent the most stated, *"The spirit-like agent makes me more relaxed during the guiding process due to its inorganic appearance"* (P16). Additionally, its appearance is consistent with the virtual environment, as another participant noted, *"The appearance is quite natural and reasonable"* (P15). However, not every participant could quickly recognize this appearance as that of a guide: *"I didn't realize the spirit-like agent was a guide when I first entered the guiding process because it is too different from guides in the real world"* (P18).

doi:10.6342/NTU202402825

4.5.5.2 Head-Only Agent

One participant who liked the head-only agent the most stated, *"The head-only agent reminds me of the interface with the circular user avatar that pops up when there's a notification"* (P18). However, most participants disliked this appearance due to the discontinuous human body. *"It is horrible that only having a head floating in the air"* (P11).

4.5.5.3 Full-Body Agent

Participants who liked the full-body agent the most noted that *"like a real guide, it's friendly"* (P3) and *"the full-body agent provides me a sense of realism; however, it makes me feel stressed"* (P5). Other participants mentioned reasons for the realism, including its humanoid appearance, body gestures, and facial expressions. However, the inconsistency between the human-like agent's appearance and behavior also caused some users to dislike it. For instance, some participants felt that the agent wasn't looking at them, or that the agent's lip and gestures weren't synchronized with the audio.

doi:10.6342/NTU202402825

Chapter 5 Discussion

Based on the virtual Dunhuang Cave 61, we developed a guiding system with an embodied conversational AI agent and further discussed aspects such as the presence, sense of presence regarding the agent's appearance. We also explored user preferences for different agent appearances and the reasons behind their choices. In this chapter, we will discuss the implications of our findings and provide suggestions for future research.

5.1 Presence and Social Presence

According to the results of the IPQ questionnaire in Figure 4.4, the agent's appearance did not cause significant differences in the user's sense of presence in the virtual environment. Combined with the positive feedback on system usability from the SUS questionnaire, we can conclude that the overall design of the tour guide system successfully integrates with the cave environment, offering good usability and effective guidance. However, we would like to discuss the differences in user interaction with different agent appearances in other aspects within the virtual space.

As we anticipated, the user study showed that the presence score for the full-body agent was significantly higher than for the other two appearances. Additionally, from the Figure 4.6, we know that the full-body agent is more noticeable to users compared

doi:10.6342/NTU202402825

to the other two appearances. As some participants mentioned, *"I found the full-body agent easier to locate during the experience."* (P1) These results align with the distraction scores, indicating that the full-body agent also caused more distraction for users than the other two appearances. As participants mentioned, *"The full-body agent would block the view of what I wanted to see."* (P14)

We found that users have different preferences for the agent's appearance at various stages of the tour. For instance, during spatial navigation, the full-body agent provides a more apparent visual guide due to its size. Additionally, during interactions with the agent, the humanoid virtual person and realistic body movements offer users a sense of realism, thereby enhancing their focus. However, when users want to view the murals on their own, the spirit-like or head-only appearances, which have a lower presence, provide a better viewing experience due to their less intrusive appearance.

In summary, social presence—where users feel as though they are being together with the agent—can indeed provide a good interactive experience. However, in a tour guide context, we need to consider additional factors, such as whether the guide can correctly lead users to hotspots or if the guide might hinder users' ability to explore the space independently.

5.2 User Preference

In the experiment, we surveyed users about their preference ranking for the three appearances and the reasons behind their choices. As shown in the Figure 4.1, more than half of the participants preferred the full-body agent the most. Through the survey, we found that many users attributed their preference for the full-body agent to its body movements

doi:10.6342/NTU202402825

(P2, P7, P16), interactive (P10, P13), and the sense of immersion it created, making them feel as if they were being guided by a real guider in the museum (P3, P11). However, why the head-only agent, which also has a human face, did not receive the same positive feedback and was even less favored than the spirit-like agent is a question we want to discuss further.

We found that even when placing the head avatar inside a bubble to reduce its abruptness, users' preference for the floating head avatar remained low. Although the head-only agent has a higher presence compared to the spirit-like agent, its discontinuous body form cause users to feel uncomfortable and find it hard to accept. During interviews, users described the head-only agent as *"creepy,"* (P11, P14) *"eerie,"*(P15) and *"unnatural."* (P5, P16) Two participants mentioned that they envisioned the agent as a *"small fairy-like figure"* (P14, P16), meaning a complete humanoid form but smaller and floating in front of the user. We think this type of agent might combine the advantages of both the full-body and head-only agents, maintaining the integrity of the human form while slightly reducing the visual obstruction.

For the spirit-like agent, a few users felt that its appearance did not intuitively suggest a guide (P9, P18). However, for most users, the appearance of the spirit-like agent resembled the image of personal voice assistants they were familiar with from their experiences, making it easy to understand. Furthermore, its technological and inorganic appearance provides users with a novel and futuristic experience.

doi:10.6342/NTU202402825

Chapter 6 Conclusion

Building on previous experiential methods for exploring the Dunhuang Caves, our main contribution is developing an interactive guide system to enhance user engagement. This system provides spatial navigation through embodied virtual agents, menus, and particle effects. Additionally, it features a conversational component powered by large language model technology, allowing users to ask the virtual agents questions actively and fostering a deeper understanding of the caves.

Our user study examined three virtual agent appearances: full-body, head-only, and spirit-like. We also investigated how facial expressions and body movements influenced interactions. The study highlights that the variety of virtual agent appearances plays a crucial role in the effectiveness of the guiding system, as different appearances can significantly shape the user experience. Our findings reveal that while body language and facial expressions enhance realism, presence, and trust, they can also be distracting. Interestingly, the different appearances did not significantly impact users' sense of presence. The full-body agent was the most preferred due to its realistic animations and interactivity, whereas the head-only appearance caused discomfort and was the least favored.

doi:10.6342/NTU202402825

Chapter 7 Limitation and Future Work

In our experiment, we found that the full-body agent significantly enhanced the realism and social presence experienced by users, making participants more focused on the conversation. This improvement in engagement and attention suggests promising potential for the full-body agent in educational contexts. However, its effectiveness in facilitating learning about the Dunhuang caves requires further verification through additional studies.

The preferred appearance of the agent varies at different stages of the tour, indicating that a hybrid appearance or one that can be hidden might be optimal. For instance, when the user is focused on viewing a mural, the agent could collapse or transform into a spirit form. During the QA phase, it could appear in a humanoid form. Even without reducing the visual obstruction caused by the agent's appearance to the user, we can optimize the agent's locomotion within the space, making the agent's positioning more spatially intelligent.

Currently, the QA system design only incorporates the user's voice input, lacking information about the user's position in the cave or their viewing details. This limitation

doi:10.6342/NTU202402825

prevents the agent from answering questions like ”What is the mural in front of me?” Therefore, providing ChatGPT with information about the user’s location and viewing details is crucial for optimizing our QA system. Integrating spatial awareness will significantly improve the agent’s ability to interact contextually with users.

Lastly, the current guiding system only works in our Dunhuang Cave 61 VR environment. To extend the system to other cultural heritage sites, we could develop a general guiding system that can adapt to different VR applications or even XR.

doi:10.6342/NTU202402825

References

[1] O. Balan, S. Cristea, G. Moise, L. Petrescu, S. Ivascu, A. D. B. Moldoveanu, F. Moldoveanu, and M. Leordeanu. ether - an assistive virtual agent for acrophobia therapy in virtual reality. In Interacción, 2020.

[2] T. W. Bickmore, L. M. Pfeifer, and D. Schulman. Relational agents improve engagement and learning in science museum visitors. In International Conference on Intelligent Virtual Agents, 2011.

[3] F. Biocca, C. Harms, and J. K. Burgoon. Toward a more robust theory and measure of social presence: Review and suggested criteria. Presence: Teleoperators & Virtual Environments, 12:456–480, 2003.

[4] C. Boletsis. The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology. Multimodal Technol. Interact., 1:24, 2017.

[5] A. Bönsch, D. Hashem, J. Ehret, and T. W. Kuhlen. Being guided or having exploratory freedom: User preferences of a virtual agent's behavior in a museum. Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents, 2021.

[6] J. B. Brooke. Sus: A 'quick and dirty' usability scale. 1996.

doi:10.6342/NTU202402825

[7] B. D. Carolis, N. Macchiarulo, and C. Valenziano. Marta: A virtual guide for the national archaeological museum of taranto. In AVI²CH, 2022.

[8] R. Chaturvedi, S. Verma, R. Das, and Y. K. Dwivedi. Social companionship with artificial intelligence: Recent trends and future avenues. Technological Forecasting and Social Change, 2023.

[9] V. Chheang, R. Marquez-Hernandez, M. Patel, D. Rajasekaran, S. Sharmin, G. Caulfield, B. Kiafar, J. Li, and R. L. Barmaki. Towards anatomy education with generative ai-based virtual assistants in immersive virtual reality environments. 2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR), pages 21–30, 2023.

[10] M. Duguleană, V.-A. Briciu, I.-A. Duduman, and O. M. Machidon. A virtual assistant for natural interactions in museums. Sustainability, 2020.

[11] V. A. R. Ecosystem. Guidebot - ar indoor navigation for museums, Sep. 2020.

[12] J. Geigel, K. S. Shitut, J. Decker, A. Doherty, and G. D. Jacobs. The digital docent: Xr storytelling for a living history museum. Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology, 2020.

[13] E. T. Hall. The hidden dimension. 1966.

[14] C. Harms and F. Biocca. Internal consistency and reliability of the networked minds social presence measure. 2006.

[15] T. Hirzle, F. Müller, F. Draxler, M. Schmitz, P. Knierim, and K. Hornbæk. When xr and ai meet - a scoping review on extended reality and artificial intelligence.

doi:10.6342/NTU202402825

Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 2023.

[16] L. Kruse, F. Mostajeran, and F. Steinicke. The influence of virtual agent visibility in virtual reality cognitive training. Proceedings of the 2023 ACM Symposium on Spatial User Interaction, 2023.

[17] J. kwang Ko, D. W. Koo, and M. S. Kim. A novel affinity enhancing method for human robot interaction - preliminary study with proactive docent avatar. 2021 21st International Conference on Control, Automation and Systems (ICCAS), pages 1007–1011, 2021.

[18] C. Kyrlitsias and D. Michael-Grigoriou. Social interaction with agents and avatars in immersive virtual environments: A survey. In Frontiers in Virtual Reality, 2022.

[19] A.-S. Milcent, A. Kadri, and S. Richir. Using facial expressiveness of a virtual agent to induce empathy in users. International Journal of Human-Computer Interaction, 38:240 – 252, 2021.

[20] N. Norouzi, K. Kim, J. Hochreiter, M. Lee, S. Daher, G. Bruder, and G. Welch. A systematic survey of 15 years of user studies published in the intelligent virtual agents conference. Proceedings of the 18th International Conference on Intelligent Virtual Agents, 2018.

[21] G. B. Petersen, A. Mottelson, and G. Makransky. Pedagogical agents in educational vr: An in the wild study. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 2021.

doi:10.6342/NTU202402825

[22] W. B. Rayward and M. Twidale. From docent to cyberdocent: Education and guidancein the virtual museum. Arch. Mus. Informatics, 13:23–53, 1999.

[23] R. Rzayev, G. Karaman, K. Wolf, N. Henze, and V. Schwind. The effect of presence and appearance of guides in virtual reality exhibitions. Proceedings of Mensch und Computer 2019, 2019.

[24] M. Saito. Effects of presentation modalities in virtual museum guides on agent impressions and painting evaluations. Proceedings of the 11th International Conference on Human-Agent Interaction, 2023.

[25] S. Schmidt, G. Bruder, and F. Steinicke. Effects of embodiment on generic and content-specific intelligent virtual agents as exhibition guides. In ICAT-EGVE, 2018.

[26] S. Schmidt, G. Bruder, and F. Steinicke. Effects of virtual agent and object representation on experiencing exhibited artifacts. Comput. Graph., 83:1–10, 2019.

[27] T. W. Schubert. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. Z. für Medienpsychologie, 15:69–71, 2003.

[28] U. Spierling, P. J. Winzer, and E. Massarczyk. Experiencing the presence of historical stories with location-based augmented reality. In International Conference on Interactive Digital Storytelling, 2017.

[29] D. Studio. Augmented reality for museums / ar guide, May. 2019.

[30] S. Sylaiou and C. A. Fidas. Virtual humans in museums and cultural heritage sites. Applied Sciences, 2022.

doi:10.6342/NTU202402825

[31] S. Sylaiou, V. Kasapakis, D. Gavalas, and E. Dzardanova. Avatars as storytellers: affective narratives in virtual museums. Personal and Ubiquitous Computing, pages 1–13, 2020.

[32] G. Trichopoulos, M. Konstantakis, G. Caridakis, A. Katifori, and M. Koukouli. Crafting a museum guide using chatgpt4. Big Data and Cognitive Computing, 2023.

[33] K. Tsitseklis, G. Stavropoulou, A. Zafeiropoulos, A. Thanou, and S. Papavassiliou. Recbot: Virtual museum navigation through a chatbot assistant and personalized recommendations. Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, 2023.

[34] L. P. Vardoulakis, L. Ring, B. Barry, C. L. Sidner, and T. W. Bickmore. Designing relational agents as long term social companions for older adults. In International Conference on Intelligent Virtual Agents, 2012.

[35] H. H. Vilhjálmsson, S. Kopp, and S. Marsella. Editorial for special issue on intelligent virtual agents. Autonomous Agents and Multi-Agent Systems, 27:197 – 199, 2013.

[36] C. Wang and Y. Zhu. A survey of museum applied research based on mobile augmented reality. Computational Intelligence and Neuroscience, 2022, 2022.

[37] I. Wang, J. Smith, and J. Ruiz. Exploring virtual agents for augmented reality. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019.

[38] L. Wanner, E. Andr, J. Blat, S. Dasiopoulou, M. Farrs, T. Fraga, E. Kamateri, F. Lingenfelser, G. Llorach, O. Martnez, G. Meditskos, S. Mille, W. Minker, L. Pragst,

doi:10.6342/NTU202402825

D. Schiller, A. W. Stam, L. Stellingwerff, F. M. Sukno, B. Vieru, and S. Vrochidis. Design of a knowledge-based agent as a social companion. Procedia Computer Science, 121:920–926, 2017.

[39] F. Weidner, G. Boettcher, S. A. Arboleda, C. Diao, L. Sinani, C. Kunert, C. Gerhardt, W. Broll, and A. Raake. A systematic review on the visualization of avatars and agents in ar & vr displayed using head-mounted displays. IEEE Transactions on Visualization and Computer Graphics, 29:2596–2606, 2023.

[40] J. W. Woodworth, N. G. Lipari, and C. W. Borst. Evaluating teacher avatar appearances in educational vr. 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pages 1235–1236, 2019.

[41] K. Zibrek, E. Kokkinara, and R. Mcdonnell. The effect of realistic appearance of virtual characters in immersive environments - does the character's personality play a role? IEEE Transactions on Visualization and Computer Graphics, 24:1681–1690, 2018.

doi:10.6342/NTU202402825

Appendix A — Questionnaire

A.1 System Usability Scale (SUS)

No	Question
1	I think that I would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.

Table A.1: SUS Questionnaire (7-point Likert scale)

doi:10.6342/NTU202402825

A.2 Igroup Presence Questionnaire (IPQ)

No	item name	Question
1	G1	In the computer generated world I had a sense of ”being there”
2	SP1	Somehow I felt that the virtual world surrounded me.
3	SP2	I felt like I was just perceiving pictures.
4	SP3	I did not feel present in the virtual space.
5	SP4	I had a sense of acting in the virtual space, rather than operating something from outside.
6	SP5	I felt present in the virtual space.
7	INV1	How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?
8	INV2	I was not aware of my real environment.
9	INV3	I still paid attention to the real environment.
10	INV4	I was completely captivated by the virtual world.
11	REAL1	How real did the virtual world seem to you?
12	REAL2	How much did your experience in the virtual environment seem consistent with your real world experience ?
13	REAL3	How real did the virtual world seem to you?
14	REAL4	The virtual world seemed more realistic than the real world.

Table A.2: IPQ Questionnaire (7-point Likert scale)

doi:10.6342/NTU202402825

A.3 User Perception of Agent

No	Question
1	How would you rate the helpfulness of the virtual agent?
2	How would you rate the presence of the virtual agent?
3	How would you rate the relatability of the virtual agent?
4	How would you rate the trust of the virtual agent?
5	How would you rate the distraction of the virtual agent?
6	How would you rate the realism of the virtual agent?

Table A.3: User Perception of Agent Questionnaire (7-point Likert scale)

A.4 Social Presence

No	Question
1	I noticed the virtual agent.
2	The virtual agent’s presence was obvious to me.
3	I was easily distracted from the virtual agent when other things were going on.
4	I remained focused on the virtual agent throughout our interaction.
5	My thoughts were clear to the virtual agent.
6	It was easy to understand the virtual agent.

Table A.4: Social Presence Questionnaire (7-point Likert scale)

doi:10.6342/NTU202402825