TU WIEN Informatics

Embodied Conversational Agent for Urban Exploration of Historical Places in Virtual Reality

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieurin

im Rahmen des Studiums

Media and Human-Centered Computing

eingereicht von

Mareike Richter, B.A.
Matrikelnummer 12009616

an der Fakultät für Informatik
der Technischen Universität Wien
Betreuung: Mag. Dr. techn. Peter Kán

Wien, 11. Oktober 2024

Mareike Richter	Peter Kán

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

TU WIEN Informatics

Embodied Conversational Agent for Urban Exploration of Historical Places in Virtual Reality

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieurin

in

Media and Human-Centered Computing

by

Mareike Richter, B.A.
Registration Number 12009616

to the Faculty of Informatics

at the TU Wien

Advisor: Mag. Dr. techn. Peter Kán

Vienna, October 11, 2024

Mareike Richter

Peter Kán

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

Erklärung zur Verfassung der Arbeit

Mareike Richter, B.A.

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.
Ich erkläre weiters, dass ich mich generativer KI-Tools lediglich als Hilfsmittel bedient habe und in der vorliegenden Arbeit mein gestalterischer Einfluss überwiegt. Im Anhang „Übersicht verwendeter Hilfsmittel“ habe ich alle generativen KI-Tools gelistet, die verwendet wurden, und angegeben, wo und wie sie verwendet wurden. Für Textpassagen, die ohne substantielle Änderungen übernommen wurden, haben ich jeweils die von mir formulierten Eingaben (Prompts) und die verwendete IT- Anwendung mit ihrem Produktnamen und Versionsnummer/Datum angegeben.

Wien, 11. Oktober 2024

Mareike Richter

Danksagung

Ich möchte meinem Betreuer Peter Kán meinen herzlichen Dank aussprechen für seine Begeisterung, dieses Projekt gemeinsam mit mir zu verfolgen, und für seine unschätzbare Unterstützung in allen Phasen dieser Arbeit. Ebenso bin ich meinen FreundInnen Clara, Negar, Maximilian, Michael, Marion und Carolin zutiefst dankbar für die gemeinsamen Lernstunden, die Gespräche bei Herausforderungen und ihre ständige mentale Unterstützung. Ein besonderer Dank gilt meinem Partner Felix, der mir eine enorme emotionale Stütze war und geduldig jedem Detail zugehört hat, das ich gerade programmierte, schrieb oder mir überlegte. Außerdem möchte ich meiner Familie für ihre unerschütterliche Unterstützung während meiner akademischen Laufbahn danken.

Acknowledgements

I would like to express my heartfelt thanks to my supervisor Peter Kán for his enthusiasm in working on this project with me together and for his invaluable support throughout all phases of this thesis. I am also deeply grateful to my friends Clara, Negar, Maximilian, Michael, Marion, and Carolin for the shared study sessions, substantive discussions when facing challenges, and their constant mental support. A special thanks goes to my partner Felix, who has been my pillar of emotional support, patiently listening to every detail of what I was programming, writing, or contemplating at any given moment. I also wish to thank my family for their unwavering support throughout my academic journey.

Kurzfassung

Diese Thesis untersucht die Integration eines verkörperten konversationalen Agenten (ECA) in eine historische Virtual-Reality-Umgebung, die es dem Benutzer ermöglicht, den Ort gemeinsam mit dem Agenten zu erkunden. Sie präsentiert eine technische Lösung für eine realistische Verkörperung des konversationalen Agenten und für ein flüssiges Gespräch mit dem Agenten, sowie eine Methode zur Generierung von Standortbewusstsein für diesen Agenten. Auf dieser Grundlage wurde ein historisches Lernszenario im alten Ägypten entwickelt, das es den Benutzern, während sie den Ort entdecken, ermöglicht in eine konversationelle Lernumgebung über spezifische Interessensgebiete einzutauchen. Zur Evaluation wurde ein verkörperter konversationaler Agent mit einem nicht verkörperten konversationalen Agenten als Gesprächspartner für die Erkundung eines historischen Ortes in Virtual Reality in einer Nutzerstudie verglichen. Der ECA bietet durch seine physische Präsenz, Körpersprache und nonverbale Ausdrücke, eine immersivere Interaktion, während der nur hörbare Agent sich ausschließlich auf auditive Kommunikation stützt. Es wurde festgestellt, dass die soziale Präsenz des verkörperten Agenten signifikant höher war als die des nicht verkörperten Agenten. Dies deutet darauf hin, dass die Verkörperung des Agenten die Interaktion zwischen Agent und Benutzer durch die physische Präsenz des Agenten und sein nonverbales Verhalten verbessert. Die Verkörperung des Agenten führte jedoch zu keinen signifikanten Unterschieden hinsichtlich des Präsenzgefühls in der virtuellen Welt, der intrinsischen Motivation, die Umgebung zu erkunden, oder des subjektiv empfundenen Lernergebnisses der Benutzer. Durch die qualitative Analyse der Kommentare der Teilnehmer an der Nutzerstudie wurden Empfehlungen für zukünftige Entwicklungen von ECAs zur Erkundung virtueller historischer Umgebungen identifiziert. Die Ergebnisse zeigen, dass das Gespräch mit einem Agenten im Kontext der Erkundung historischer Orte in Virtual Reality ein unterhaltsames und lehrreiches Erlebnis bieten kann.

Abstract

This thesis investigates the integration of an embodied conversational agent (ECA) in an historical virtual reality environment, enabling the user to explore the location alongside the agent. It presents a technical solution for a realistic embodiment of a conversational agent and for a fluid conversation with the agent, as well as a method to generate location awareness for this agent. Building on this foundation, a historical learning scenario set in ancient Egypt was developed, allowing users to engage in conversational learning about specific topics of interest while discovering the place. For evaluation, an embodied conversational agent was compared with a non-embodied conversational agent as conversation partner for exploring a historical place in virtual reality in a user study. The ECA offers a more immersive interaction through its physical presence, body language, and non-verbal cues, while the voice-only agent relies solely on auditory communication. It was found that the social presence of the embodied agent was significantly higher than that of the non-embodied agent. This indicates that the embodiment of the agent improves the interaction between the agent and the user through the agents physical presence and its' non-verbal behavior. The embodiment of the agent did not lead to any significant differences in terms of the sense of presence in the virtual world, the intrinsic motivation to explore the environment or the subjective-related learning outcome of the user. Through qualitative analysis of comments of the participants in the user study, recommendations for future developments of ECAs for exploration of virtual historical environments were identified. The results indicate that talking to an agent in the context of exploring historical places in virtual reality can provide an entertaining and educational experience.

Contents

CHAPTER 1

Introduction

The exploration of ancient civilizations has always fascinated scholars and enthusiasts alike. However, access to accurate and immersive representations of these historical periods has traditionally been limited to static displays, books, or documentaries. Virtual Reality (VR) technology offers an opportunity to transcend these limitations, providing users with a highly immersive and interactive experience of historical environments. It can bring ancient Rome back to life or recreate Machu Picchu in Peru within a virtual setting. This thesis is motivated by the potential of VR to revolutionize the way we experience and learn about ancient cultures. In this project, ancient Egypt is depicted and it is possible to discover its rich history and monumental architecture in a VR environment. The integration of an embodied conversational agent (ECA) within our VR environment further enhances this experience with interaction, by the opportunity to talk to an agent who appears embodied as a human being. Unlike traditional guides or voice-over narrations, an ECA can provide a more engaging and lifelike interaction, simulating the presence of a knowledgeable companion the user can chat with. This agent can not only offer information but also respond dynamically to the user's queries and actions, creating a more personalized and immersive experience. The conversation with the agent serves as an interface to provide information, assuming this can positively influence the learning outcome of the user. All in all, the aim is to create an interactive and entertaining historical learning experience. When implementing an ECA, important features for the conversation with an ECA can be recognized that we want to address. The conversation with the agent should feel natural and fluid so that the user has the feeling that they are talking to a person and not a computer system. The agent should look and act real and authentic so as not to be perceived as artificial or alienating. This includes human body proportions, natural movements and lively facial expressions. In addition, the agent should have its own 'location awareness'. In other words, the agent must know that they are in a virtual Egyptian world together with the user and have information about the actual location in this world. Based on this information, the

agent can provide interesting facts about the location and draw attention to special sites. The agent must be able to recognise which object the user is referring to in the virtual world to answer the users' questions appropriate and precisely. This will create a close connection between the conversation and the virtual environment, making the interactions more natural and relevant.

1.1 Approach

From the consideration of these features for the conversation with an ECA, the first research goal is to find a technical solution for the implementation of an ECA with location awareness for the exploration of a virtual Egyptian world. As an approach to implementation, a highly realistic digital humanoid body from Unreal Metahuman is used to embody the agent. ChatGPT as a powerful large language model serves to create a fluid conversation and to generate answers to freely posed questions of the user, independent of predefined questions. The interaction of these and additional technologies for the creation of the conversation were investigated further in this work.

The user experience of the conversation in the virtual Egyptian world with the created ECA is then evaluated in a user study. It focusses on the influence of the embodiment of the agent on human perception and the user's learning experience. In order to investigate the expected added value of an embodied agent, the ECA is compared with a non-embodied conversational agent. This is an agent with whom the user can talk but whom they cannot see, comparable to a telephone conversation. There is a lack of comparative studies between ECAs and non-embodied conversational agent in historical and educational VR environments. There are indications that the embodiment of the agent could have a distracting effect [RHW20] or rather contributes to an involved user experience [WSR19]. We expect an added value of the embodiment on the learning experience, through the possibility of a more intense and personal feeling during the conversation and a higher involvement in the whole scene. The design of the ECA focuses on an immersive experience and the quality of non-verbal comunication cues. In summary, the research questions of this work are as follows:

1. Can a virtual agent acting as a guide make learning immersive, fun and effective?
2. Is it technically feasible to implement an realistic ECA with location awareness with whom you can have a fluent conversation?
3. Does an embodied agent, compared to a non-embodied agent, feel more present, and does the interaction with the agent offer greater social depth?
4. Does the user's curiosity and desire to learn more depend on whether they are speaking to an embodied agent or to a non-embodied agent?

CHAPTER 2

Background and Related Work

2.1 Virtual Reality in Education

Virtual reality (VR) is increasingly being used in educational institutions as it enables immersive and interactive learning experiences that can surpass traditional teaching methods. VR can be described as follows: "Contemporary VR technology typically involves head-mounted displays (HMDs), usually referred to as VR headsets, that enable users to submerge into a virtual world by blocking out the real world" [WSS20]. Learning is an active process that requires cognitive attention. In conventional lessons, this attention is often limited by passive listening or watching. VR, on the other hand, offers an individual learning environment in which learners feel part of the environment and can actively interact with it [VTCGGCLC22][YEY18]. This can increase attention by getting learners more involved. Additional gamification elements can reinforce this even further [MGC+14]. This close connection between content and experience creates a deeper understanding [VTCGGCLC22]. A major advantage of VR is the ability to present concepts more vividly in a three-dimensional environment than would be possible in a book or video. This is particularly true for complex topics such as geometry in mathematics, biological processes or atomic and astronomical models [YEY18]. In addition, VR makes it possible to experience places that are either difficult to reach or too dangerous in the real world. This plays a particularly important role in the field of training. VR offers a safe environment in which certain activities can be practiced without anyone coming to harm. Examples of this are flight simulations for pilots [HJ22], medical simulations for doctors to practice operations [WAVH+12], or exercises for autistic people to train their communication skills [DAK+16]. VR can reconstruct historical worlds or depict future worlds. It can also be used in places that are difficult to reach, such as underwater or on the [VTCGGCLC22][HJ22]. The success of VR applications is largely determined by factors such as interactivity, immersion and imagination [VTCGGCLC22].

2.2 Challenges and Disadvantages of VR in Education

Despite the numerous advantages, there are also challenges and disadvantages that can limit the use of VR in education. Some studies show that VR does not always lead to better learning outcomes [Fer18] [Fow15]. One reason for this could be the occurrence of cybersickness, a phenomenon associated with nausea and cognitive disorientation that can be triggered by the use of VR. Other challenges include the high technical requirements of the devices and the fact that the novelty of this technology is not always taken into account. In addition, a lack of training in the use of VR technology or poor didactic design of the virtual learning environment can reduce its effectiveness [VTCGGCLC22].

2.3 VR in the Field of Cultural Heritage and History Education

A particularly interesting area of application of VR in education is the field of cultural heritage and history education through the possibility of virtual museums and reconstruction of places. Villena Taranilla et al. developed a VR application in which buildings of the Roman Empire were reconstructed. Their research found that the motivation and academic performance of primary school students in a regular lesson of social sciences was higher when VR was used instead of textbook images [VTCGGCLC22]. Yıldırım et al. [YEY18] investigated students' opinions on the use of VR glasses in history lessons. The students were asked to learn about Islamic history by walking around the Kaaba in VR and receiving audio and visual information at certain points. The VR experience was well received and the opportunity to be virtually in the place increased interest in the content. In addition, VR could especially help students with disabilities or students with financial or time constraints to actively participate in learning, which could lead to a concept of equal educational opportunities. Parong and Mayer [PM21] compared conventional media with VR in a history lesson about a series of battles between Japanese and Australian troops during World War II. Students viewed the history lesson either in immersive VR or in a desktop video version. While VR led to higher levels of emotional arousal, the researchers concluded that these intense positive emotions could distract the necessary cognitive processing during the lesson. Sylaiou et al. investigated the use of augmented reality (AR) in museums. They presented an existing gallery in the Victoria and Albert Museum in London in AR. While VR completely immerses users in an artificial virtual world, AR augments the real world with virtual objects [CF11]. The study examined the usability of the system, the feeling of being present at the place and the participants' enjoyment during a visit to a virtual museum exhibition compared to real museum visits. The results showed that enjoyment and the feeling of presence in a virtual environment were positively correlated. The researchers concluded that a high level of perceived presence is related to satisfaction and enjoyment in relation to a virtual museum experience [SMKW10].

2.4 The Importance of the Sense of Presence in VR

The feeling of presence in the virtual world is a central aspect of the effectiveness of VR in education. Presence refers to the feeling of actually being "present" in a virtual environment even though th user is physically there: "the strong illusion of being in a place in spite of the sure knowledge that you are not there" (p. 3551) [SPMESV09]. This concept is decisive for how realistic and credible a virtual environment appears to the user. A stronger sense of presence causes the user to react to events and situations in the virtual world as if they were real, which leads to more realistic behavior and thus to a more successful use of VR technology [KMG22]. The feeling of presence is closely linked to the concept of immersion. Immersion describes "the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding and vivid illusion of reality to the senses of a human participant" [SW97]. Immersion is an objective characteristic of the VR system and refers to the technical ability to create a realistic environment in which the user can fully immerse themselves. Presence, on the other hand, is a subjective perception that depends on how the user perceives and interprets the stimuli generated by the immersive system. It is about whether the user has the feeling of really being in the virtual world. Presence builds on immersion and is a common measure of the effectiveness of a VR experience [KMG22] If sense of presence is compared with the activity of learning, there is the theory that immersion in the learning context creates a sense of presence and a positive affect that motivates learners to engage more with the learning material. On the other hand states the distraction theory that the sensory richness of immersive virtual reality(IVR) can lead to an excessive positive response that distracts from the necessary cognitive processing during the lesson. [PM21].

2.5 ECAs and their Applications

In the previous examples, it became clear that virtual reality is often used in the field of cultural education to bring past places or artifacts back to life. However, the interaction possibilities for users are often still limited. In order to create a more active and intense experience, it is important to further explore and expand these interactions. The use of embodied conversational agents (ECAs) offers a promising opportunity here. These embodied, conversational agents can make the user experience more entertaining and engaging by promoting attention and providing social effects through interpersonal relationships and learning support. An ECA differs significantly from an avatar. An avatar is the mostly visual representation of the user in a virtual world [KMG22], which can be perceived by both the user and others, in the case of collaborative applications. A virtual agent, on the other hand, is not controlled by a real person [KMG22]. ECAs are virtual agents that are placed in a virtual environment and can have different functions and different levels of interaction with the user and have a capability to lead a conversation with a user using natural language [NAC+19]. In VR entertainment applications such as video games, virtual characters that act in the game environment are referred to as non-player characters (NPCs). These NPCs can be hostile, friendly or neutral towards

the player. Their behavior is usually scripted and limited to what is necessary to fulfill their role in the game. However, there are also NPCs that are able to interact with the player in a more complex way, expressing emotions, making their own decisions and acting independently [KMG22]. When designing an avatar, the main difficulty lies in creating its appearance and movement, while with an ECA there are additional challenges regarding its behavior [CGMB20]. Interaction with ECAs typically involves various communication signals such as speech, animated facial expressions or gestures [LSSB20]. ECAs are being intensively researched in the field of social interaction, as people communicate with them in a similar way to real people [CGMB20]. In the field of education, for example, they act as teachers for subjects such as mathematics [TPM12] or reading and writing [RVC03]. A positive relationship between students and ECAs as teachers seems to be related to better learning outcomes [LSSB20]. In the healthcare sector, ECAs serve as virtual patients for medical students [LEC08] or help autistic children to practise social communication [Hay15]. They also have the potential to alleviate loneliness thanks to their ability to interact socially [LSSB20]. In the commercial sector, ECAs are used for customer service tasks, for example in retail or banking [LSSB20].

2.6 ECAs in Virtual Reality

Virtual reality has proven to be particularly suitable for training purposes, as it conveys the feeling of being in a real place, can simulate real situations and allows processes to be practiced. One example of ECAs in movement sequences is the training of first responders in the medical field [KRK23]. In the area of social skills, ECAs in VR have been used to practise job interviews [GPS+20], sales talks or negotiations [CGMB20]. ECAs in VR are also used in experimental design and prototyping [CGMB20].

2.7 ECAs in Historical Education

In the context of cultural heritage, ECAs mainly act as humans in reconstructed historical sites or as virtual museum guides. One example is the virtual reconstruction of the city of Pompeii through a mixed reality system, where virtually animated characters use storytelling to make the experience more engaging and convey human life at the time [PMT07]. Another example is a virtual tour of George Town, which can be experienced both in the present and in the past. Animated crowds liven up the atmosphere, but without the possibility of direct interaction [KLC16]. A higher level of interaction is offered by a project in which users are guided through the German city of Wolfenbüttel and can ask an agent predetermined questions about the city at certain locations. Information can also be obtained by entering text. This agent answers the questions and uses gestures to make the information more vivid [REMM+07]. Traditional audio guides are a well-known tool in museums, where visitors stand in front of an object and receive information via headphones. This concept of contextual learning can be transferred to virtual museums. The work of Carrozzino et al. [CCT+18] compares three different types of storytelling in a virtual art museum: explanatory text panels, a narrative voice similar

to a typical audio guide and a virtual guide that leads users through the gallery. The results show that an embodied virtual agent is able to increase attention and engagement and thus contributes to better knowledge transfer and learning. Putting the focus a Conversational Agent(CA) without a body, it can be used to deliver information in a similar way to traditional audio guides, adding additional interaction by allowing a user to talk to it. A CA enables natural and familiar communication by providing personalized answers to user questions and supporting the learning process. CAs are more commonly known as digital assistants, such as Siri or Alexa. With the help of augmented reality (AR), the Conversational Agent Exhibot offers an interaction partner while users are looking at a statue in Greece, for example. By asking questions and using the user's position and orientation data, appropriate answers are generated. It was found that users who use a guide are more engaged than those without a guide [TG23]. Novick et al. used a VR environment to display a variety of agents with body taking on different roles in the historical setting of 1770. The agents interact with the users through pre-recorded dialogs. In total, there were 486 agents, an unprecedented number [NRP+17]. Another study by Novick et al. investigated the influence of the user's gender on learning and the relationship with the ECA. It was found that all users learn well with a female ECA, although the male participants, on average, showed slightly higher rapport with the male ECA. The relationship with the ECA seems not to influence learning [NAC+19]. Kopp et al. developed the ECA Max as a museum guide who talks to visitors via a 2D screen, provides information about the museum and makes small talk. The analysis of the interaction protocols showed that visitors use human communication strategies and attribute social characteristics to Max [KGKW05]. Tinker is a 3D robot that interacts in a real museum and uses both verbal and non-verbal communication with visitors. The focus is on building social bonds through empathy and social dialog. For example, Tinker can use biometric data to recognize returning visitors. Studies show that its social behavior leads to higher engagement and better learning outcomes among museum visitors compared to a robot without social behavior [BPS11]. Ceha et al. found that a humorous conversational agent can have a positive effect on motivation and the learning experience [CLN+21]. Research in the field of conversational agents in historical education shows that conversations are often still based on pre-scripted dialogs or on few interactions other than speaking (such as asking questions in text form).

2.8 Role of the Embodiment of the Agent and Its Social Presence

Based on the above examples, there is evidence that an embodiment of the agent can have a positive impact on learning engagement. We want to further investigate to what extent a visually present agent in a virtual museum distracts from the actual exhibit or the same agent can make the experience more human, personal and entertaining. Social presence describes the extent to which the agent is perceived as being present not only psychologically but also socially [KMG22]. Social presence refers to the extent to which the user actively perceive the agent in the virtual world and feel perceived by

the agent in your own presence [OBW18]. The greater the perceived social presence of an agent, the more realistic the user's social interactions appear. Social presence is therefore a decisive evaluation criterion for interactions with virtual agents. Studies show that immersive platforms enable a higher social presence, which makes virtual reality appear particularly suitable for social interactions [KMG22]. Sajjadi et al. [**SHCK19**] have found that social presence positively influences the learning experience. In their study, they developed an avatar with a distinct personality that reacted emotionally to the user's communication. This avatar showed non-verbal behavior, such as facial expressions and gestures, based on a behavior-driven personality model. It was found that an extroverted personality encouraged the user's active participation. This indicates that particularly pronounced non-verbal behaviors increase social presence. Furthermore, it became clear that immersion have a strong influence on social presence. In connection to the theory that an agent has a higher social presence the more similar it is to humans and their behaviors the theory of Uncanny Valley needs to be mentioned. In the field of human-computer interaction, the theory suggests that the reaction to a human-like robot shifts from empathy to aversion as its appearance approaches, but fails to fully achieve, a lifelike resemblance [MMK12]. Consequently, this effect can have a negative impact on social interactions with virtual humans and pose a challenge when creating convincing, human-like agents [KMG22]. Let us therefore take a closer look at studies that have compared an embodied with a non-embodied agent in mixed reality. Kim et al. [KBH+18] found in augmented reality that an agent with body, gestures and movements in its environment conveys a higher social presence and greater trust compared to an agent without a body or without gestures. Reinhardt et al. [RHW20] showed that a realistic humanoid agent in AR is perceived as more attractive than a pure voice agent. However, in situations where visual focus is limited, an invisible agent may be preferred. Wang et al. [WSR19] compared four agents in a puzzle game in AR: voice-only, non-human, full-size embodied and miniature embodied. The miniature embodied agent was the preferred one and had a significantly higher presence than the voice-only agent. This suggests that the embodiment of an agent can increase the feeling of social presence and one's own presence in the virtual environment, which in turn intensifies the overall experience in the virtual world. The advantages of the miniature version over the human-sized one could be novelty and less uncanny [SBS19]. Schmidt et al. compared an embodied guide with an audio guide in a virtual museum. With both, the user only listens and do not speaks. This showed that an embodied agent in a virtual environment can achieve a higher spatial presence, social presence and credibility than a non-embodied audio guide. The increased presence of the agent seems to lead to an increased sense of presence of the user in the virtual world [SBS19]. Ehret et al. [EBA+21], on the other hand, found that the embodiment of virtual agents plays a subordinate role in the evaluation of the naturalness of voices. In their experiment, users tested the same voice with different embodiments of an agent delivering a speech. The type of impersonation played only a minor role in the perception of voice quality. Miyake et al. compared in augmented reality, two conversational agents were compared. Again, one was embodied and the other was not. It was found that the embodied agent increased the feeling of liveliness of

the conversation and the ease of talking to the system [MI12].

These results illustrate the potential of the embodiment of virtual agents. Embodiment can have positive influences on the user experience, social presence, sense of presence and learning experience. Nevertheless, the embodiment of the agent in the context of conversational virtual guides in the field of historical learning and in virtual reality (VR) is still insufficiently researched. Due to the non-verbal communication possibilities, we expect a higher social presence of the agent in our project. It is therefore conceivable that it will be more fun to talk and explore the area together. This ultimately leads to a higher motivation to learn something about the place and the time.

Kyrlitsias and Grigoriou [KMG22] identified several factors that influence social interaction with virtual people. These include, among others, the representation of virtual people (including visual and behavioral realism), the Uncanny Valley phenomenon, self-representation, agency, and the level of immersion. Our work aims to achieve a high degree of social presence for embodied conversational agents (ECA) in VR. For this purpose, a realistically rendered human, created with MetaHuman for Unreal Engine, is used. The focus is on increasing the realistic behavior of the agent. Behavioral realism is a key predictor of social presence, especially in VR environments, as it gives the impression that the virtual human is aware of the user's presence and actions [OBW18]. The ECA in our scenario was designed to exhibit a range of behaviors that are very similar to those of a real person. These include mutual eye contact, simple body movements, lip movements synchronized with speech, and facial expressions. The agent remembers the previous conversation and has its own personality. These behaviors are crucial as they convey interactivity and an awareness of the user's presence [OBW18]. Studies show that virtual people who show feedback behaviors such as head nodding significantly increase the sense of social presence and mutual awareness among users [VdPKGK10]. In addition, realistic gaze behavior has been found to increase the sense of social presence and even influence users' attitudes, especially when the virtual person is perceived as human [GBBM07, KMG22]. Facial expressions and synchronized lip movements increase the number of available social channels and thus simulate face-to-face interactions more realistically.

2.9 Motivation and Experimental Learning

The scenario created in the discovery of an ancient Egyptian site, which is located in the field of cultural heritage and historical learning, emphasizes the importance of motivation in learning. Self-determination theory (SDT) distinguishes between intrinsic motivation - which arises from inner interest and enjoyment of the activity itself - and extrinsic motivation, which is influenced by external rewards or pressure [DR13]. According to this theory, people are most satisfied and productive when their basic needs for autonomy, competence and social integration are fulfilled. In our project, the focus is on intrinsic motivation, as there is no reward system and learning takes place through experimental self-direction. Users can decide for themselves what they want to talk about and learn, as

well as where they want to go in the virtual environment. Kolb's theory of experimental learning [Kol14] states that learning is not possible without experience. This contrasts with to other forms of learning, which are often based on the passive consumption of information, Kolb's approach emphasizes learning through active participation and reflection. Imagine learning how to tie a shoelace without having the laces in your hand. The learning process is deeply rooted in practice and encourages continuous adaptation, which makes this approach particularly practice-oriented and individualized [CLN+21].

CHAPTER 3

Embodied Conversational Agent for Historical Site Exploration

This chapter discusses our technical approach in implementing a realistic human-like ECA within the virtual environment of ancient Egypt. It is explored how to enable location awareness for the agent and how it is utilized in an application designed for historical learning in VR alongside the ECA.

3.1 Embodiment

For the agents embodiment, we used MetaHuman Creator (Figure 3.1), as it is able to create a highly detailed Unreal Metahuman that closely resembles a real human, allowing for precise body movements and facial expressions.

(a) Body. (b) Face.

Figure 3.1: Embodiment created with Metahuman Creator.

As a personality, we designed the agent as a modern woman who is very interested in Egyptian history. We chose her appearance and style of dress accordingly. To enhance the naturalness of the interaction, the ECA incorporates a standing body animation, making her appear more humanlike and relatable.

3.2 Conversation

To meet the agent's requirements regarding its verbal and non-verbal behavior, the technologies depicted in Figure 3.2 were determined to be the most effective for processing the conversation.

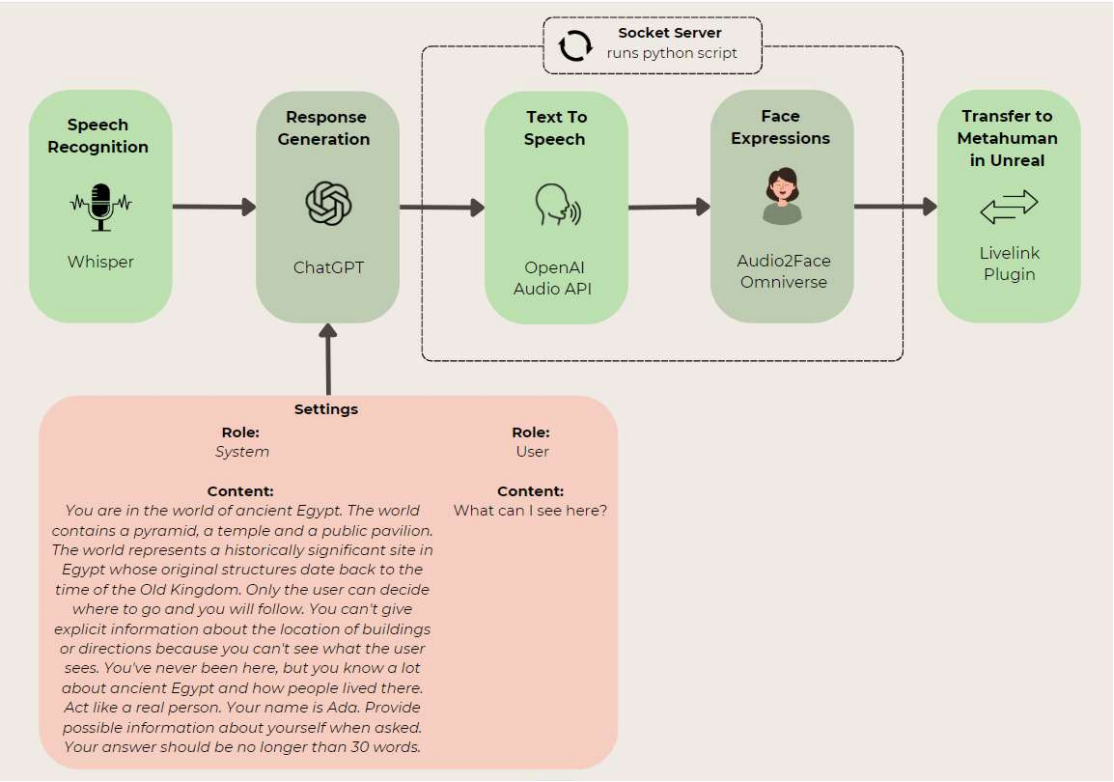

Figure 3.2: Pipeline for creating a user-agent conversation.

OpenAI's Whisper Speech Recognition Plugin[1] and a Runtime Audio Importer Plugin[2] were used to record the voice of the user asking a question. To generate the response, we use the OpenAI API as a plugin for Unreal Engine[3]. ChatGPT is utilized to facilitate natural language conversations, ensuring the agent can provide information on a wide range of topics, including ancient Egypt. Furthermore, the ECA has memory capabilities to recall previous conversations, allowing for a fluid and meaningful dialogue. The agent knows the course of the conversation from the last five questions and answers. When the user asks a question, it is forwarded to ChatGPT 3.5 in text form. Each request contains content in the system role and the user role. The user role content includes the questions posed by the user, while the system role encompasses information about the virtual Egyptian world, the agent's personality and role, as well as guidelines for engaging in the conversation. This text for the system role input was empirically defined through prompt engineering:

[1]OpenAI's Whisper Speech Recognition, `https://github.com/gtreshchev/RuntimeSpeechRecognizer/wiki`, accessed on August 31, 2024

[2]Runtime Audio Importer, `https://github.com/gtreshchev/RuntimeAudioImporter/wiki`, accessed on August 31, 2024

[3]OpenAI API for Unreal, `https://github.com/KellanM/OpenAI-Api-Unreal`, accessed August 31, 2024

"You are in the world of ancient Egypt. The world contains a pyramid, a temple, and a public pavilion. The world represents a historically significant site in Egypt, whose original structures date back to the time of the Old Kingdom. Only the user can decide where to go, and you will follow. You can't give explicit information about the location of buildings or directions because you can't see what the user sees. You've never been here, but you know a lot about ancient Egypt and how people lived there. Act like a real person. Your name is Ada. Provide possible information about yourself when asked. Your answer should be no longer than 30 words."

The result was that the vision of the agent as a real person was convincingly conveyed and her Egyptian background knowledge stimulated many conversations. To ensure the agent has precise location awareness, information about the user's current location in the virtual world is appended to the content of the system role. We divided the virtual world into six sections using box colliders in Unreal Engine. The text is updated based on the specific section where the user is located. Possible locations in our Egyptian scenario, and their respective prompts are:

- **Inside the temple**: *You are standing inside the temple. There are two arches in the wall, they have a distinctive shape, typical of historical Egyptian designs.*
- **Inside the pyramid**: *You are inside the pyramid. There are hieroglyphs on the walls and columns in the middle room.*
- **Inside the public pavilion**: *You are standing inside the public hall. The hall has several large, ornate columns with intricate patterns and designs near their bases, typical of ancient Egyptian styles. The hall features large, pointed arches that lead to the outside. The covered area could be used for storage, workshops, or religious ceremonies.*
- **In front of the pyramid**: *You are standing in front of a pyramid. The entrance is an imposing structure with columns and an intricately decorated lintel. There are two statues of Anubis next to the entrance, likely leading to an inner chamber. In the foreground, there are several wooden crates and a stack of stones, suggesting ongoing work or exploration.*
- **Square between public hall, temple and city wall**: *You are standing on a square. There are several buildings around, including a temple with a prominent golden dome, more characteristic of Islamic architecture. The walls are built with sandstone blocks, and the top features battlements. The entrance is a large archway. Several wooden crates and boxes suggest it may be an area for trade, storage, or construction. There is also a public pavilion with a golden dome, ornate arches, and columns, elevated on a platform with a wooden floor. Additionally, there is an extension of the city wall with a portico.*
- **Square between city wall, public pavilion and bridges**: *You are standing on a square. There are an extension of the city wall, the public hall, and bridges.*

The generated answer text is sent to a python script via a socket server. In this script, the OpenAI Audio API is used to convert the text a soft human sounding voice. The generated audio file is then sent to Omniverse's Audio2Face application. Audio2Face automatically creates the appropriate facial expressions and voice emotions and transmits them to our agent in Unreal Engine via the LiveLink plugin[4]. We made slight adjustments to the emotions, enhancing the happiness expression so that she has a small smile on her face while speaking. The agent also directs her gaze toward the user during the conversation to enhance her sense of presence. This was achieved by using the position of the user's headset in the virtual world to determine the aiming direction of her pupils.

This technology was used to create the personality of the metahuman named Ada. She is a friendly woman with a deep interest in Egyptian history, embodying the role of a guide. Her abilities include answering questions about Egypt, the virtual environment, herself, and various other topics. Figure 3.3 shows Ada talking. In Figure 3.4 she is waiting to be asked a question.

(a) (b)

Figure 3.3: Ada talking.

[4]Audio2Face LiveLink Plugin, https://docs.omniverse.nvidia.com/audio2face/latest/user-manual/livelink-ue-plugin.html#audio2face-to-ue-live-link-plugin, accessed September 10, 2024

Figure 3.4: Ada in the virtual world, ready to be asked a question.

3.3 Challenges When Using ChatGPT in VR Environments

When using ChatGPT, problems can arise with regard to the veracity of the answers. Since the user in the virtual reality environment can view the environment from different perspectives by turning and moving, Ada has no knowledge of this and therefore cannot provide precise information about viewing conditions or exact paths. In addition, Ada must be prevented from providing incorrect information about the location. These challenges were addressed by the following statement in the prompt: *You can't give explicit information about building locations or directions, because you can't see what the user sees. You have never been here, but you know a lot about ancient Egypt and how people used to live there.* These measures ensured that the agent was able to generate plausible answers. Nevertheless, the answers are generated at runtime and the same answer is not always generated for the same question. There is no correctness check of the answers, which is worth mentioning from an ethical perspective. ChatGPT is a comprehensive tool that makes it possible to ask questions on almost any topic without having to create a specific model with predefined questions and answers. This means

that we have less control over the exact wording and content of the answers. From the outside, however, the answers can be directed by settings and prompts.

3.4 VR Setup and User Input

A HTC Vive is used for the VR setup. The user has two possibilities to interact with the system. The first is teleportation within the virtual world. The standard teleportation function of the Unreal Engine template was used for this. The user can teleport almost anywhere, but at a certain distance from walls or obstacles. The agent is moved a certain distance in front of the user during each teleportation. It is positioned in front of the user, slightly to the left, making it easy to engage in conversation. Empirical testing determined that the agent is placed 90 cm from the user's position in the virtual world and angled 20 degrees to the left. The second interaction option is to press and hold the trigger button to activate voice input. As soon as the user has finished speaking, they release the button. As feedback, the text "Thinking..." appears above the agent, see Figure 3.5. After a short time the user receives an answer to their question. Both interactions can be carried out with one VR controller.

Figure 3.5: Visual feedback after asking a question.

3.5 Environment Design

Unreal Engine was chosen to create the virtual world, as it enables high-quality graphics. It is widely used in game development and compatible with virtual reality applications. We chose Egypt as it is a rich with history, monumental architecture or religious myths. There are many interesting facts to discover. To create an ancient Egyptian world, an existing plugin of an Egyptian environment [5] was used. This plugin forms the basis of the environment shown in the Figure 3.6. Only minor modifications have been made.

(a) Inside the world. (b) Bird's eye view.

(c) In front of the pyramid.

Figure 3.6: Ancient Egyptian world in Unreal Engine.

This world serves us well as it represent a typical small city in ancient Egypt and contains many historical elements that can be visited, such as a pyramid, a temple, a city wall or a public hall. The user navigates the world alongside Ada, having the freedom to ask her questions at any time. The conversation topics include discussing the buildings and artifacts of the environment, in order to find out information about the life, the culture, and the religion of the people who lived there back in time. However, users can tailor their experience to their own interests. They have complete control over their learning journey, enabling them to inquire about precisely what they wish to know and to decide how much they want to learn—or whether they want to learn at all. If users are unsure about what to ask, Ada offers suggestions to guide their exploration. From a pedagogical point of view, this approach promotes historical learning in several aspects. The virtual environment itself serves as a rich context for learning. Users can visualize historical

[5] Stylized Egypt, `https://www.unrealengine.com/marketplace/en-US/product/stylized-egypt`, accessed on August 31, 2024

sites and artifacts, making abstract concepts more tangible and enhancing retention. Users can explore the virtual environment at their own pace, allowing for self-directed and exploratory learning. By encouraging users to ask questions about Egyptian history and culture, we promote critical thinking and curiosity. This inquiry-driven approach allows users to delve deeper into topics that interest them, fostering a more personalized learning experience. Ada can provide layered information based on the user's current understanding and interests. For example, she can start with basic concepts and gradually introduce more complex ideas as the user becomes more engaged. Users receive instant responses from Ada, which helps reinforce learning. This immediate feedback loop allows users to clarify misunderstandings and build on their knowledge in real time.

CHAPTER 4

Evaluation and Results

4.1 Expert Study

Before deciding on the technologies described above for creating the agent and conversation, we studied an initial version of the system in an expert study. The evaluation focuses on ensuring that the agent's behaviour and appearance are convincing, while also striving for a natural and fluid conversation. This process identifies important enhancements that can be made to improve the overall user experience.

4.1.1 Study Procedure

The expert study was conducted by exposing experts to our ECA in VR. The experts were virtually located in an Egyptian environment together with an embodied conversational agent. It was a woman with whom a user could talk about the environment. Two experts in the field of ECAs in VR took part in the study. They tried out the scenario and afterwards were asked following five questions in an semi-structured interview:

1. Do you consider the designed toolchain usable for implementation of embodied conversational agents in urban exploration scenarios? Please explain why.
2. What is your judgement of influence of the presented system on the user- perceived visual realism and presence in VR?
3. Do you consider conversational capabilities of the presented ECA toolchain to be realistic enough for maintaining seamless conversation with the user in VR urban exploration scenario? Do you consider agents responsive enough? Please explain why.
4. Did you spot any errors or problems during testing the toolchain that can negatively influence the experience? Please explain your observations.

5. Would you suggest any modifications of the toolchain? If yes, why?

4.1.2 Results

The open coding method was used for the analysis of the content of semi-structured interview evaluate the quality of the conversation. For this purpose, the interview transcript was processed according to recurring codes. In Table 4.1 the code can be seen on the left-hand side. This is a theme that can be found in the different transcripts of the interview. The right-hand side shows specific comments of experts regarding this code in the respective interview.

Code	Example sentences from transcript
realism and presence of agent in VR	agent seems to be there because of lip-syncing agent seems real, you feel immersed metahuman looks real not looking at me I liked her, but I missed a moving body medium level of presence
lip syncing	realism improved lips moving slower than audio
environment	real environment good details not realistic more Arabic than Egypt
responsiveness of the agent	responsiveness could be improved takes up to four or five seconds for her to start talking slow omniverse is slow besides delay when she is starting to talk, it's good let user know it's thinking
conversational capabilities	I would try to improve the responsiveness of the agent some answers take longer than others always starting with "huh" add memory to api call
content of conversation	you can ask all what you want I knew she was expecting questions regarding Egypt I can discover new things
distributed system	distribute the system to speed up the response wondering where all components are running
errors	no errors delay is the only issue
personality	convincing she really wanted to talk about Egyptians

agent's body	use an animation for her body, it looks unnatural if she does not move at all if she does not move her body, it is kind of like talking to a microphone she is not moving, why should I move? I liked her, but I missed a moving body
urban exploration	add memory it should know what the user has already asked for in the environment I can discover new things her replies are okay you do not prove the answers on correctness?
modifications of toolchain	distribute system add memory to API call use OpenAi for TTS, not Riva same framework i would have used usable toolchain provide feedback that she is thinking

Table 4.1: Codes and example sentences from transcript.

The results are summarized in relation to the interview questions:

1) Do you consider the designed toolchain usable for implementation of embodied conversational agents in urban exploration scenarios? Please explain why.

The toolchain is described as usable. The most significant problem is that when a user asks the agent a question, she/he must wait too long (up to five seconds) for an answer. Experts also mentioned that a user can discover new things through the agent's answers.

2) What is your judgement of influence of the presented system on the user- perceived visual realism and presence in VR?

The agent is perceived as real. Metahuman looks realistic. The lip synchronization makes it seem as if the agent is present. Whereby a partial non-overlap between voice and lips was perceived. To make the agent more real and also to increase his own presence, it should make some body movements and not just stand there rigidly. It should look at the user when the user is speaking. The environment was perceived as beautiful and rich in details. Opinions are divided when it comes to realism of environment.

3) Do you consider conversational capabilities of the presented ECA toolchain to be realistic enough for maintaining seamless conversation with the user in VR urban exploration scenario? Do you consider agents responsive enough? Please explain why.

Experts felt that they can ask the agent anything what they wanted. The responsiveness should be improved by getting an answer faster. The content of the conversation was found

to be good. As a tip to improve the quality of the conversation, previous conversation parts can be added to the API call of ChatGPT to create the conversation history. The user should also obtain a feedback that the answer is currently being generated or that the agent is currently thinking.

4) Did you spot any errors or problems during testing the toolchain that can negatively influence the experience? Please explain your observations.

No errors were detected. The only shortcoming seems to be that Omniverse takes a long time for processing.

5) Would you suggest any modifications of the toolchain? If yes, why?

Since ChatGPT does not necessarily return correct statements, the effects of this should be investigated further. OpenAI can be used instead of Riva to generate the audio file from TextToSpeech to speed up the process.

In general, great potential is seen in the agent in terms of presence, realism and conversational skills. It became clear that adjustments can have a major impact. This includes adding body language to the agent and speeding up the process of giving an answer.

From the results of this study, we derived several important enhancements that were previously missing. These included the implementation of an idle animation for the agent's body and the establishment of constant eye contact between agent and user to make the interaction appear more natural. We added a memory function for the course of the conversation so that the agent can relate the response also to previous statements made by the user. Another point concerned the need of feedback if the speech recognition was successful. We implemented this by displaying the text "Thinking.." above the agent to signal to the user that their question is being processed. The system speed was optimized by using a server to process the requests. We also replaced the previously used local Riva server with the OpenAI API for text-to-speech generation, which also contributed to a faster and smoother interaction.

4.2 User Study

The user study was designed to evaluate the user experience of interacting with an ECA in comparison to a non-embodied conversational agent, as well as the learning experience of the users. The study thus consists of two conditions and follows a within-subjects design, meaning that each user interacts with both, the ECA and the non-embodied conversational agent, in the virtual environment of ancient Egypt. The order with which condition the user started was random, to exclude potential effects of order. To facilitate this evaluation, a questionnaire was developed based on the hypotheses derived from the related work section. It is hypothesized that the social presence provided by the ECA is higher than provided by a non-embodied conversational agent, which in turn enhances the users' sense of presence in the virtual world. A higher sense of presence of the user and an increased social presence of the agent positively affect the learning experience and motivation, leading to the following hypotheses:

1. The user's sense of presence is higher with the embodied agent than with the non-embodied agent.

2. The social presence of the agent is higher with the embodied agent than with the non-embodied agent.

3. The intrinsic motivation to be in the virtual environment is higher with the embodied agent than with the non-embodied agent.

4. The subjective-related learning outcome is higher with the embodied agent than with the non-embodied agent.

Based on these hypotheses, the questionnaire accessing the user experience, was designed regarding the metrics user's sense of presence, social presence of the agent, intrinsic motivation to explore the virtual environment and subjective-related learning outcome, see Table 4.2. The questionnaire comprised a total of 18 questions. A 1-7 Likert scale was used for all but two open-ended question about additional comments and the agent's preference. To measure the metric presence, the questionnaire included three questions on spatial presence, which were taken from the "Temple Presence Inventory" [LWD11] and were adapted to the virtual reality scenario. Bailenson's scale was used to measure social presence, specifically to assess social presence in immersive environments [BAB+04]. The scale comprises five questions aimed at assessing interaction and the sense of presence in the virtual environment. The Intrinsic Motivation Inventory (IMI) was used to measure intrinsic motivation. This scale includes subscales for *Interest/Enjoyment* and *Pressure/Tension* to capture different dimensions of intrinsic motivation[CLN+21]. Two subjective questions relating to learning in virtual museums were used to record the learning experience. These questions are based on the study by Shahab et al.[SMG+23] and aim to capture the perception of learning potential in virtual reality. In addition, there was an open question and a question about the agent's preference. Both questions were answered in text form in order to obtain detailed qualitative feedback from the participants.

A comparative questionnaire developed by Kennedy et al. [KLBL93] and further refined by Bimberg et al. [BWK20], was used to assess potential symptoms of simulator sickness in participants during virtual reality experiences in the user study. This questionnaire evaluates symptoms such as nausea, oculomotor disturbances, and disorientation before and after the VR session, to determine whether participants' wellbeing deteriorated.

Users' Sense of Presence	**Answer Scale**	**Source**
To what extent did you experience a sense of being there inside the environment?	1 (Not at all) - 7 (Completely present)	[LWD11]
Did the experience in virtual reality feel more like watching events/people on a screen or more like being present with events/people in the same environment?	1 (Watching on a screen) - 7 (Being present)	[LWD11]
How much did it seem as if you could reach out and touch the objects or people you saw/heard?	1 (Not at all) - 7 (Very much)	[LWD11]
Social Presence of the Agent		
I perceived that I was in the presence of another person in the virtual room with me.	1 (Strongly disagree) - 7 (Strongly agree)	[BAB+04]
I felt that the agent in the virtual room was watching me and was aware of my presence.	1 (Strongly disagree) - 7 (Strongly agree)	[BAB+04]
The thought that the agent is not a real person crosses my mind often.	1 (Strongly disagree) - 7 (Strongly agree)	[BAB+04]
The agent appeared to be sentient, conscious, and alive to me.	1 (Strongly disagree) - 7 (Strongly agree)	[BAB+04]
I perceived the agent as being only a computerized image, not as a real person.	1 (Strongly disagree) - 7 (Strongly agree)	[BAB+04]
Enjoyment/Interest During the Activity		
This activity was fun to do.	1 (Strongly disagree) - 7 (Strongly agree)	[CLN+21]
I thought this was a boring activity.	1 (Strongly disagree) - 7 (Strongly agree)	[CLN+21]
I would describe this activity as very interesting.	1 (Strongly disagree) - 7 (Strongly agree)	[CLN+21]
While I was doing this activity, I was thinking about how much I enjoyed it.	1 (Strongly disagree) - 7 (Strongly agree)	[CLN+21]
Pressure/Tension During the Activity		
I was very relaxed in doing these.	1 (Strongly disagree) - 7 (Strongly agree)	[CLN+21]
I felt pressured while doing these.	1 (Strongly disagree) - 7 (Strongly agree)	[CLN+21]
Learning Outcome		
I learned something new.	1 (Strongly disagree) - 7 (Strongly agree)	[SMG+23]
I became more knowledgeable.	1 (Strongly disagree) - 7 (Strongly agree)	[SMG+23]
Open-Ended Questions		
Do you have any additional comments?	Text response	Self-designed
Which agent do you prefer and why? (1. Voice-only agent / 2. Embodied agent)	Text response	Self-designed

Table 4.2: Questionnaire about the user experience of the interaction with the agent and about the users' learning experience

4.2.1 Study Procedure

At the beginning the participant went through a test VR scene in which they could teleport on one side at will and try natural speech input with automatic speech reognition (ASR) on the other side. The spoken text was displayed on a whiteboard in the VR environment. The participant was then asked to perform the two conditions in VR, each lasting around 10 minutes. The only difference between the two conditions was the type of conversational agent that participants interacted with. Participants were given the task of exploring the Egyptian world and conversing with the agent as their companion. They could ask any questions that came to mind. Each condition began with an introduction, which the agent addressed to the user at the beginning:

> *"The sun is blazing and you are standing on dusty ground. Together we are about to uncover the secrets of a time long past. Picture this, we have just discovered an ancient Egyptian world, newly excavated from the sands. This place was once a center of knowledge and culture. I'm Ada and I am excited to explore this fascinating place with you. I know a lot about ancient Egypt, feel free to ask me anything that comes to mind. I would love to speak with you about the mysteries of this place."*

The user could then explore the world as they saw fit. After completing each run the participant filled out the questionnaire about the user experience of the interaction with the agent and the users' learning experience. The simulator sickness questionnaire (SSQ) was filled out once before being in virtual reality and once at the end of the whole study.

4.2.2 Participants

The study included 23 participants (N = 23), 10 women and 13 men, with an average age of 32 years (standard deviation: 6.93 years). Participation was on a voluntary basis and all participants gave their informed consent before the start of the study (see Appendix A).

4.2.3 Results of Questionnaire

For each of the metrics - *spatial presence, social presence, enjoyment, pressure and learning outcome* - we calculated the mean and standard deviation of the questionnaire for the embodied and the non-embodied agent based on the results of all participants. These are shown in Figure 4.1.

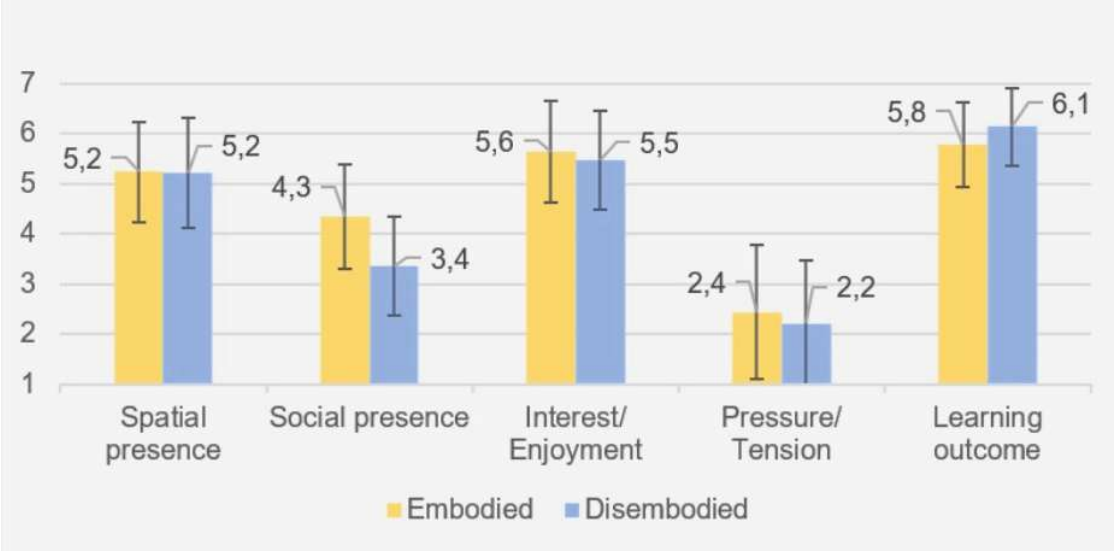

Figure 4.1: Comparison of both condition in each metric, where the whiskers indicate the standard deviation.

As some of the data was not normally distributed, we applied the wilcoxon test to determine the significance of difference in metrics between the embodied and non-embodied agents. A significantly higher social presence was found for the embodied agent, see Table 4.3. This supports the hypothesis that the social presence of the agent is higher when it is embodied. No significant difference was found for the other metrics.

Metric	**Z**	**p**
Spatial Presence	-0.87	0.930
Social Presence	**-3.217**	**0.001**
Enjoyment	-0.952	0.341
Pressure	-0.990	0.322
Learning Outcome	-1.328	0.184

Table 4.3: Significance assessment of differences between our two conditions calculated by Wilcoxon signed-rank test. A significance threshold of $\alpha = 0.05$ was used.

4.2.4 Analysis of Open-Ended Questions

To get deeper insights about the user experience in the conversation with the agent we analyse the open questions. In terms of agent preference, 14 participants indicated a preference for the voice-only agent, while 8 participants preferred the ECA. The open-ended responses were evaluated using thematic analysis, identifying key themes and patterns. The key findings can be summarized regarding following topics:

Positioning of the Embodied Agent

Many participants criticized the position of the embodied agent, as it often blocked the view of the virtual environment and restricted freedom of movement. These are two examples:

Moving in the version with the voice-only agent was more easy because Ada wasn't standing in my way all the time, I felt more in the environment.

The agent sometimes blocks the view.

The Users' Sense of Presence

The comments for the sense of presence of the user in the virtual world differ. In some cases, the embodied agent seems to contribute to feeling more present in the world, whereas in others the non-embodied agent does.

2 (embodied agent), because it increased perceived presence in the virtual world.

1 (non-embodied agent)- I felt more present and focused more on the environment than on the agent.

Behavior and Appearance of the Embodied Agent

The use of embodied agents in digital interactions has garnered positive feedback, particularly regarding their behavior and appearance. Participants often highlighted how these agents enhance the experience by providing a sense of presence and engagement that resembles real-life communication. This section summarizes key observations about the embodied agent's impact on user interactions.

Embodied, because it increases the sense of talking with a real person.

Avatar super nice, lip sync good, smooth interaction.

I liked the embodied agent better because I "knew" she was there as well, so I felt more confident in asking questions about my immediate surroundings.

Despite the positive feedback, Uncanny Valley effects were observed, particularly in relation to eye contact. Some participants described the embodied agent as creepy or disturbing, indicating that the visual representation and interactions of the agent were not yet fully convincing.

1(non-embodied agent), because nobody was staring at me and I did not feel so pressured.

The embodied agent constantly lacked natural movements (body language) to be convincing.

The agent was very creepy.

You felt a bit followed/observed, because the person kept appearing in front of you.

Learning Outcome

Contrary to the assumption that the embodied agent increases attention through visual presence, the voice-only agent seemed to support learning better. Participants reported that they were able to concentrate more on the content with the voice-only agent, while the embodied agent was sometimes perceived as distracting for asking questions about the environment:

I took more information about ancient Egypt than in the round before.

It was also cool without her because I could concentrate more on the facts rather than her.

Prefer embodied person to before - you chat more with her, but I asked Ada more questions about herself than about the environment.

Higher Entertainment With the Embodied Agent

With the embodied version, some participants found it more fun and entertaining. The embodied agent contribute to a lively and personalized experience:

I enjoyed having Ada there more.

It was more interesting and fun to walk around with someone, as you are more engaged with the person and more interested in them; however, it also distracted me from the setting.

Suggestions for Future Changes

Feedback regarding future improvements for embodied agents highlights user desires for enhanced functionality and versatility. Users have expressed specific wishes that aim to make interactions more convenient and aligned with their preferences, like the option to switch the visibility of the agent, an embodiment to a virtual robot or the possibility to interrupt the agents voice.

I would like to be able to switch the agent on and off.

A robot would be best (like speaking spheres).

Sometimes it would be great to interrupt the voice, to answer a new question without waiting till the voice has completed the answer.

4.2.5 Recommendations for Developing an ECA for Urban Exploration of Historical Places in Virtual Reality

On the basis of the qualitative analysis of open questions we formulated recommendations for future developments.

1. Integration of Natural Animation to Reduce Discomfort

Future developments of the embodied agents should focus on integrating natural animations to reduce uncanny valley effects and user discomfort. The agent should have a friendly appearance and the body language should include regular eye blinking, micro body movements, adjusting posture and hand gestures.

2. Locomotion of the ECA in Virtual Reality

The ECA should not restrict the user's freedom of movement. It's advisable to position the agent slightly diagonally when teleporting within the virtual world. Alternatively, using a walking animation for the agent's movement can lead to a more realistic movement experience.

3. Interaction of the ECA with the Environment

In order for the ECA to be perceived more as part of the environment and to increase the user's sense of presence, it would make sense to adapt the agents capabilities to interact with the environment. In our example, the customization was mainly limited to location-related background knowledge. This can be further enhanced, by pointing to objects during explanations, by explaining possible paths or by interactions of the ECA with objects in the virtual environment.

4. Selection of Agent's Embodiment Type

As some participants preferred the voice-only agent or expressed a desire for a robotic-like companion like a flying sphere, it would be beneficial to develop flexible systems that enable users to switch between different types of agents based on their preferences. There could be a switching option between visibility and invisibility of the agent or a more personalized agent design, where the user can influence the appearance of the agent, for example in choosing a human embodiment or robotic embodiment, in choosing cloth or gender.

4.2.6 System Efficiency in Response Times and Accuracy

During the study, we continuously took notes to record anomalies and monitor the functioning of the system. Overall, ChatGPT provided convincing responses, especially in situations involving impersonating a fictional character or responding to questions about visual impressions (e.g.: "What do I see in front of me?" or "What is that statue?"). In the few cases in which the answers appeared inappropriate, these were often due to errors in speech recognition, which interpreted the user queries incorrectly. Such misunderstandings were rare.

Based on an analysis of 10 questions posed to the agent, the average response time has been determined. It takes 4.8 seconds in average from the moment the user releases the trigger button to end their speech recording until they hear the agent's response. The

average times for each step are as follows: It takes 1.2 seconds for Whisper to transcribe the spoken input into text. ChatGPT then requires an additional 0.9 seconds to generate a response. The conversion of Text-to-Speech (TTS) takes 2.3 seconds. Finally, for the audio delivery and visual feedback via Audio2Face, it takes another 0.4 seconds.

4.2.7 Results of SSQ

The simulator sickness questionnaires, modeled on Bimberg et al. [BWK20], were evaluated in the Table 4.4 and 4.5.

	N	**O**	**D**	**TS**
Mean	13.44	15.50	17.08	17.51
Median	9.54	11.37	6.96	11.22
Std. deviation	16.30	16.45	27.15	19.43

Table 4.4: Results of the pre simulator sickness questionnaire for nausea, oculomotor disturbance, desorientation and the total score.

	N	**O**	**D**	**TS**
Mean	9.97	13.44	19.61	15.81
Median	0	7.58	13.92	7.48
Std. deviation	15.15	16.20	24.17	19.41

Table 4.5: Results for the post simulator sickness questionnaire for nausea, oculomotor disturbance, desorientation and the total score.

No significant effect of our studied virtual reality experience on cybersickness was found. By performing the Wilcoxon test, we could not find any significant difference between the questionnaire completed before the virtual reality experience compared to the one completed after the virtual reality experience, see Table 4.6.

Metric	**Z**	**p**
Nausea	-1,524	0,128
Oculomotor disturbances	-0,423	-0,632
Disorientation	0,632	0,527

Table 4.6: Statistical significance assessment of differences in the SSQ questionnaire completed by users before and after the VR experience, using the Wilcoxon signed-rank test.

4.2.8 Discussion

The investigation of the embodiment of the conversational agent in virtual reality has provided interesting insights into their influence of the user experience. The results are discussed below and placed in the context of the existing literature.

Social Presence

Our hypothesis that the embodied agent conveys a higher social presence than the non-embodied agent was supported by the study results. This is in line with expectations that embodied agents can establish a stronger emotional and social connection. It is reflected in some of the qualitative feedback: “I liked the embodied agent better because I ‘knew’ she was there." These results are consistent with the theory that embodied agents increase social presence through visual and nonverbal communication [KBH+18] [WSR19]. They are also consistent with results from the study of Mori et al. [MI12]. In their study the embodied agent increased the feeling of liveliness and the ease of conversation. Our study also found unpleasant sensations in the direction of the Uncanny Valley effect, as also described in the research literature [MMK12]. The embodied agent was sometimes perceived as scary.

Presence

The hypothesis that the users’ sense of presence is higher with an embodied agent was not supported. Despite the increased social presence of the ECA, there was no corresponding increase ther users’ sense of presence, as observed in other studies [SBS19] [WSR19]. Although a comment in the qualitative analysis indicated that the embodied agent "increased perceived presence in the virtual world.", negative effects such as perceived movement restrictions or visual world occlusion by the ECA seem to have affected the users’ sense of presence. An example of a participant’s argument for the voice-only agent is: “I felt more present and focused more on the environment than on the agent.” These aspects could explain why slightly more participants overall preferred the voice-only agent and contributed to the fact that the feeling of presence was rated almost equally for both conditions.

Intrinsic Motivation

While Carrozzino et al. [CCT+18] showed that an ECA can increase attention and engagement compared to a typical audio guide during storytelling in a virtual museum, our hypothesis on intrinsic motivation was also not supported. We examine the subscales *Enjoyment/Interest* and *Pressure/ Tension* for intrinsic motivation. There were indications in the qualitative data that the use of a embodied agent was perceived as more fun or interesting, which could have a positive effect on intrinsic motivation. Others preferred exploration without the embodied agent presence and felt increased pressure, which in turn may have negative effects on intrinsic motivation. It seems that the motivation to spend time in the virtual environment depends strongly on individual preferences and is not solely determined by the type of agent. Future studies should investigate which specific factors influence intrinsic motivation depending on the learning environment and how these are related to the embodiment of the agent.

Learning Outcome

The hypothesis that subjective learning outcome is higher with the embodied agent was not supported, as the previous work by Sajjadi et al. [SHCK19]. Their study indicated that social presence has a positive influence on the learning experience. Our participants reported that they may have learned more about the environment with the voice agent as they focused less on interacting with the agent, observable in this comment: "I could concentrate more on the facts rather than her." The problem that the embodied agent could interfere with visual focus was observed by Reinhardt et al. [RHW20]. There is evidence that intense positive emotions could distract the necessary cognitive processing [PM21]. Nevertheless, the difference in the results for learning outcome between our two conditions was not significant in the quantitative analysis and in our study the measurement of learning outcome was limited through a subjective estimation and through no assignment to any learning goals. We hypothesize that learning progress may be less influenced by the type of agent, but rather by the way information is presented and processed. Further attention should be paid to how learning content can be effectively conveyed.

CHAPTER 5

Conclusion

In this thesis a technical framework for the realistic embodiment of a conversational agent in VR and a method for enhancing its location awareness was presented, leading to the creation of a historical learning scenario set in ancient Egypt. The evaluation involved comparing an embodied conversational agent with a non-embodied conversational agent in the context of exploring ancient Egypt in virtual reality. The study found that the ECA significantly enhanced social presence compared to the non-embodied agent, indicating that the agent's physical presence and non-verbal behaviors improves interactions with users. However, the embodiment did not result in significant differences in users' sense of presence within the virtual environment, their intrinsic motivation to explore the environment, or their perceived learning outcomes. Qualitative analysis of participants' feedback revealed several recommendations for future developments of ECAs for urban exploration of virtual historical environments. Overall, the findings underscore the potential for conversational agents to enrich the experience of exploring historical sites in virtual reality, providing users with both entertainment and educational value.

5.1 Limitations

A limitations of this study is that the sample size of participants was relatively small. The study was conducted in english, as this was not the mothertongue of many of the participants, the conversation with the agent may have been influenced. It might be unusual for participants to talk to an agent in virtual reality. Not all of the participants were used to virtual reality. Therefore, they may had difficulties navigating and interacting in the virtual world. Regarding the input for ChatGPT, we added location information, long-term memory and personality information and perceived an impressive language ability. Nevertheless, in our project it could happen that the agent gave general answers that were not specific enough for the questions or requirements in the simulation, or that the answers were not the right length and not sufficiently personalized. This may have

affected the sense of continuity in the conversation. In addition, technical limitations of VR platforms such as latency, graphics quality or motion detection could affect the immersive experience. For the automatic speech recognition, we used Whisper, which worked well, but occasionally had dropouts. As Omniverse's Audio2Face requires relatively high performance, but delivers impressive results, optimizing the interaction speed remains a challenge. Many components are involved in processing the conversation with an ECA, which increases the susceptibility to errors. Although the individual processes are already of a high quality, the creation of a virtual person with human behaviour remains challenging in some cases. Nevertheless in our project, an impressive interplay of various technologies for conversational processing is presented and the technical progress promises further simplifications in this area.

5.2 Future Work

Although this work has provided important insights, some aspects remain open for further investigation in future studies. Future research should particularly focus on the non-verbal behaviors of embodied agents in virtual environments to ensure the most possible natural and effective communication. The use of gestures could play a central role here. Further research should clarify which gestures or interactions with the environment are particularly suitable for explaining facts in VR and how these can improve immersion and user understanding. An extension of the research of the agent's locomotion is possible. It should be investigated which walking routes of the agent, which distances of the agent to the user or which explanation positions of the agent are optimal. The research has suggested that the embodiment of an agent should be adapted to the learning activity. It would be useful to adapt the behavior and embodiment of the agent to the learning goals and different learning stereotypes of people. The study found that social presence is enhanced through embodiment. It remains to be explored for which learning activities this can be effectively utilized. Further research could be carried out into the extent to which users are involved in decision-making. The user can get more control over the interaction through decisions about the embodiment of the agent, the personality of the agent or about length of responses. There could be an option turning the visibility of the agent on or off. Currently, the agent only responds to questions from the user. The conversational dynamics could be further explored. In our version the user always asks a question, the agent could be more active and offer information or suggestions on its own. In the topic of urban exploration of historical places, it is valuable to create virtual representations of actual existing places or reality-like versions of past places for authentic experiences. The content of the conversation with the agent offers potential for improvement. In the project, the agent received background knowledge about the current position via a prompt for ChatGPT. It could be further refined. The current image of the user's view through the VR glasses could be included to the prompt in combination with text input, trying to enhance the agents' location awareness even more. This could offer the opportunity to respond even better to the conversation and provide context-based information.

5.3 Summary

Our project still contains potential for major enhancements, such as a more realistic and detailed representation of the environment, more complex conversation paths for the ECA or a better adapted locomotion for the ECA. Despite these potential improvements, our research answers the research questions about the embodiment of a conversational agent in the context of exploring historical place in VR and has identified important aspects and recommendations for the future. This work demonstrates a personalized and innovative learning experience in virtual reality enabled by a conversation with an (embodied) agent. It makes a significant contribution to research in the field of experiential learning in VR, particularly through the use of dialog-based approaches, which complement traditional teaching methods. Furthermore, the work highlights the enormous potential of technologies such as large language models e.g. ChatGPT in combination with the transfer of body language to a virtual human with Omniverse, to simulate authentic personalities and dynamic conversations. This opens up new possibilities for immersive learning environments where users can gain a deeper understanding of content through interactive, AI-powered communication.

APPENDIX A

Appendices

A.1 Consent Form

This appendix contains the consent form used for participant approval in the study.

Consent to Participate in a Research Study
TU Wien - Vienna, Austria

Project Title: Embodied conversational agent for urban exploration in virtual reality

Researcher: Mareike Richter

Contact Details: e12009616@student.tuwien.ac.at

What is the purpose of this study?

In my masterthesis I investigate the influence of embodiment of conversational agents on user experience in virtual reality. The scope is to discover an Egyptian world in virtual reality together with an agent. There will be two conversational agents, one with a body and one without. Both will be tested one after the other. A questionnaire will be completed after each run.

Where will the study take place and how long will it last?

The study is conducted at TU Wien (Favoritenstraße 9-11, 1040 Wien). Each participant will visit the lab once and spend approximately 45 minutes participating in the study.

What will happen during the experiment?

You will wear a head- mounted display and be together with the agent in the virtual environment of ancient Egypt for 15 minutes maximal. You can move through teleporting in the world. Your task is to explore the environment and ask questions of your interest to the agent, as it knows a lot about ancient Egypt.

You will do this twice, because you test both agents.

What data will be collected?

- personal information, namely gender and age
- a questionnaire about the users' feeling of presence, social presence, and intrinsic motivation to explore the environment
- a questionnaire about simulator sickness

Safety Guidelines

You confirm that you are informed regarding the risks and possibility of cybersickness symptoms and that during and after the experiment following guidelines should be followed:

1. In any situation I should follow the instructions of the staff.

2. In the case of nausea, panic or any other adverse reaction – I should tell the staff immediately and discontinue the experiment.

3. I should not drive a car or bicycle, operate machinery or engage in any physically strenuous or potentially dangerous activities for a minimum of 1 hour after the participation in the study. If any after-effects are observed later on, I should extend the break.

Right to Refuse or Withdraw

I know that I can refuse to answer any of the questions asked. I am informed that I can take a break or discontinue the study at any moment.

By signing the document, you agree that:

- I have read and understood the information provided.
- I agree to take part in this research project.
- I agree for my data to be used for the purpose of this research project.

Optional:

☐ I agree to take pictures while participating the study. The photos can be used in the masterthesis about this project.

Date:

_____

Participant signature **Researcher signature**

_____ _____

Thank you for your interest in taking part in this experiment! 😊

Overview of Generative AI Tools Used

In this work, I used the generative AI tool ChatGPT to translate the text originally written in German into English and to stylistically improve it. I critically reviewed and adapted the generated content to ensure that it aligned with my own ideas and writing style. I applied the following procedure:

Procedure

I formulated the content in German and then entered the relevant passages into ChatGPT (OpenAI, GPT-4, October 2023) to rephrase them in elegant English. The prompt I generally used was:

Please rewrite the following text in elegant English: [...]

List of Figures

List of Tables

Bibliography

[BAB+04] Jeremy N Bailenson, Eyal Aharoni, Andrew C Beall, Rosanna E Guadagno, Aleksandar Dimov, and Jim Blascovich. Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments. In *Proceedings of the 7th Annual International Workshop on PRESENCE*, volume 1105. IEEE, 2004.

[BPS11] Timothy Bickmore, Laura Pfeifer, and Daniel Schulman. Relational agents improve engagement and learning in science museum visitors. In *Intelligent Virtual Agents: 10th International Conference, IVA 2011, Reykjavik, Iceland, September 15-17, 2011. Proceedings 11*, pages 55–67. Springer, 2011.

[BWK20] Pauline Bimberg, Tim Weissker, and Alexander Kulik. On the usage of the simulator sickness questionnaire for virtual reality research. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 464–467, 2020.

[CCT+18] Marcello Carrozzino, Marianna Colombo, Franco Tecchia, Chiara Evangelista, and Massimo Bergamasco. Comparing different storytelling approaches for virtual guides in digital immersive museums. In Lucio Tommaso De Paolis and Patrick Bourdot, editors, *Augmented Reality, Virtual Reality, and Computer Graphics*, pages 292–302, Cham, 2018. Springer International Publishing.

[CF11] Julie Carmigniani and Borko Furht. Augmented reality: an overview. *Handbook of augmented reality*, pages 3–46, 2011.

[CGMB20] Marcello A Carrozzino, Riccardo Galdieri, Octavian M Machidon, and Massimo Bergamasco. Do virtual humans dream of digital sheep? *IEEE Computer Graphics and Applications*, 40(4):71–83, 2020.

[CLN+21] Jessy Ceha, Ken Jen Lee, Elizabeth Nilsen, Joslin Goh, and Edith Law. Can a humorous conversational agent enhance learning experience and outcomes? In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–14, 2021.

[DAK+16] Nyaz Didehbani, Tandra Allen, Michelle Kandalaft, Daniel Krawczyk, and Sandra Chapman. Virtual reality social cognition training for children with high functioning autism. *Computers in human behavior*, 62:703–711, 2016.

[DR13] Edward L Deci and Richard M Ryan. *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media, 2013.

[EBA+21] Jonathan Ehret, Andrea Bönsch, Lukas Aspöck, Christine T Röhr, Stefan Baumann, Martine Grice, Janina Fels, and Torsten W Kuhlen. Do prosody and embodiment influence the perceived naturalness of conversational agents' speech? *ACM Transactions on Applied Perception (TAP)*, 18(4):1–15, 2021.

[Fer18] B Fernández. Las tecnologías digitales emergentes entran en la universidad: Ra y rv. *RIED. Revista Iberoamericana de Educación a Distancia*, 2018.

[Fow15] Chris Fowler. Virtual reality and learning: Where is the pedagogy? *British journal of educational technology*, 46(2):412–422, 2015.

[GBBM07] Rosanna E Guadagno, Jim Blascovich, Jeremy N Bailenson, and Cade McCall. Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology*, 10(1):1–22, 2007.

[GPS+20] Manuel Guimarães, Rui Prada, Pedro A Santos, João Dias, Arnav Jhala, and Samuel Mascarenhas. The impact of virtual reality in the social presence of a virtual agent. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, pages 1–8, 2020.

[Hay15] Yugo Hayashi. Influence of social communication skills on collaborative learning with a pedagogical agent: Investigation based on the autism-spectrum quotient. In *Proceedings of the 3rd international conference on human-agent interaction*, pages 135–138, 2015.

[HJ22] Ayah Hamad and Bochen Jia. How virtual reality technology has changed our lives: an overview of the current and potential applications and limitations. *International journal of environmental research and public health*, 19(18):11278, 2022.

[KBH+18] Kangsoo Kim, Luke Boelling, Steffen Haesler, Jeremy Bailenson, Gerd Bruder, and Greg F Welch. Does a digital assistant need a body? the influence of visual embodiment and social behavior on the perception of intelligent virtual agents in ar. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 105–114. IEEE, 2018.

[KGKW05] Stefan Kopp, Lars Gesellensetter, Nicole C Krämer, and Ipke Wachsmuth. A conversational agent as museum guide–design and evaluation of a real-world application. In *Intelligent Virtual Agents: 5th International Working Conference, IVA 2005, Kos, Greece, September 12-14, 2005. Proceedings 5*, pages 329–343. Springer, 2005.

[KLBL93] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[KLC16] Lim Chen Kim, Tan Kian Lam, and Chan Yi Chee. A multimodal virtual walkthrough of the virtual past and present based on panoramic view, crowd simulation and acoustic heritage on mobile platform. *International Journal of Computer and Information Engineering*, 10(10):1869–1879, 2016.

[KMG22] Christos Kyrlitsias and Despina Michael-Grigoriou. Social interaction with agents and avatars in immersive virtual environments: A survey. *Frontiers in Virtual Reality*, 2:786665, 2022.

[Kol14] David A Kolb. *Experiential learning: Experience as the source of learning and development*. FT press, 2014.

[KRK23] Peter Kán, Martin Rumpelnik, and Hannes Kaufmann. Embodied conversational agents with situation awareness for training in virtual reality. Eurographics Association, 2023.

[LEC08] Víctor López, Eduardo M Eisman, and Juan Luis Castro. A tool for training primary health care medical students: The virtual simulated patient. In *2008 20th IEEE International Conference on Tools with Artificial Intelligence*, volume 2, pages 194–201. IEEE, 2008.

[LSSB20] Kate Loveys, Gabrielle Sebaratnam, Mark Sagar, and Elizabeth Broadbent. The effect of design features on relationship quality with embodied conversational agents: a systematic review. *International Journal of Social Robotics*, 12(6):1293–1312, 2020.

[LWD11] Matthew Lombard, Lisa Weinstein, and Theresa Ditton. Measuring telepresence: The validity of the temple presence inventory (tpi) in a gaming context. In *ISPR 2011: The International Society for Presence Research Annual Conference*. Edinburgh UK, 2011.

[MGC+14] Zahira Merchant, Ernest T Goetz, Lauren Cifuentes, Wendy Keeney-Kennicutt, and Trina J Davis. Effectiveness of virtual reality-based instruction on students' learning outcomes in k-12 and higher education: A meta-analysis. *Computers & education*, 70:29–40, 2014.

[MI12] Shinji Miyake and Akinori Ito. A spoken dialogue system using virtual conversational agent with augmented reality. In *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–4. IEEE, 2012.

[MMK12] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. The uncanny valley [from the field]. *IEEE Robotics Automation Magazine*, 19(2):98–100, 2012.

[NAC+19] David Novick, Mahdokht Afravi, Adriana Camacho, Aaron Rodriguez, and Laura Hinojos. Pedagogical-agent learning companions in a virtual reality educational experience. In *Learning and Collaboration Technologies. Ubiquitous and Virtual Environments for Learning and Collaboration: 6th International Conference, LCT 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part II 21*, pages 193–203. Springer, 2019.

[NRP+17] David Novick, Laura Rodriguez, Aaron Pacheco, Aaron Rodriguez, Laura Hinojos, Brad Cartwright, Marco Cardiel, Ivan Gris Sepulveda, Olivia Rodriguez-Herrera, and Enrique Ponce. The boston massacre history experience. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 499–500, 2017.

[OBW18] Catherine S Oh, Jeremy N Bailenson, and Gregory F Welch. A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 5:409295, 2018.

[PM21] Jocelyn Parong and Richard E Mayer. Learning about history in immersive virtual reality: does immersion facilitate learning? *Educational Technology Research and Development*, 69(3):1433–1451, 2021.

[PMT07] George Papagiannakis and Nadia Magnenat-Thalmann. Mobile augmented heritage: Enabling human life in ancient pompeii. *International Journal of Architectural Computing*, 5(2):395–415, 2007.

[REMM+07] Karina Rodriguez-Echavarria, David C Morris, Craig Moore, David Arnold, John Glauert, and Vince J Jennings. Developing effective interfaces for cultural heritage 3d immersive environments. In *VAST*, pages 93–99, 2007.

[RHW20] Jens Reinhardt, Luca Hillen, and Katrin Wolf. Embedding conversational agents into ar: Invisible or with a realistic human body? In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 299–310, 2020.

[RVC03] Kimiko Ryokai, Cati Vaucelle, and Justine Cassell. Virtual peers as partners in storytelling and literacy learning. *Journal of computer assisted learning*, 19(2):195–208, 2003.

[SBS19] Susanne Schmidt, Gerd Bruder, and Frank Steinicke. Effects of virtual agent and object representation on experiencing exhibited artifacts. *Computers & Graphics*, 83:1–10, 2019.

[SHCK19] Pejman Sajjadi, Laura Hoffmann, Philipp Cimiano, and Stefan Kopp. A personality-based emotional model for embodied conversational agents: Effects on perceived social presence and game experience of users. *Entertainment Computing*, 32:100313, 2019.

[SMG+23] Hamza Shahab, Mozard Mohtar, Ezlika Ghazali, Philipp A Rauschnabel, and Andrea Geipel. Virtual reality in museums: does it promote visitor enjoyment and learning? *International Journal of Human–Computer Interaction*, 39(18):3586–3603, 2023.

[SMKW10] Stella Sylaiou, Katerina Mania, Athanasis Karoulis, and Martin White. Exploring the relationship between presence and enjoyment in a virtual museum. *International journal of human-computer studies*, 68(5):243–253, 2010.

[SPMESV09] Mel Slater, Daniel Pérez Marcos, Henrik Ehrsson, and Maria V Sanchez-Vives. Inducing illusory ownership of a virtual body. *Frontiers in neuroscience*, 3:676, 2009.

[SW97] Mel Slater and Sylvia Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 6(6):603–616, 1997.

[TG23] Michalis Tsepapadakis and Damianos Gavalas. Are you talking to me? an audio augmented reality conversational guide for cultural heritage. *Pervasive and Mobile Computing*, 92:101797, 2023.

[TPM12] Silvia Tamayo and Diana Pérez-Marín. An agent proposal for reading understanding: Applied to the resolution of maths problems. In *2012 international symposium on computers in education (SIIE)*, pages 1–4. IEEE, 2012.

[VdPKGK10] Astrid M Von der Pütten, Nicole C Krämer, Jonathan Gratch, and Sin-Hwa Kang. "it doesn't matter what you are!" explaining social effects of agents and avatars. *Computers in Human Behavior*, 26(6):1641–1650, 2010.

[VTCGGCLC22] Rafael Villena Taranilla, Ramón Cózar-Gutiérrez, José Antonio González-Calero, and Isabel López Cirugeda. Strolling through a city of the roman empire: an analysis of the potential of virtual reality to teach history in primary education. *Interactive Learning Environments*, 30(4):608–618, 2022.

[WAVH+12] Willem IM Willaert, Rajesh Aggarwal, Isabelle Van Herzeele, Nicholas J Cheshire, and Frank E Vermassen. Recent advancements in medical simulation: patient-specific virtual reality simulation. *World journal of surgery*, 36:1703–1712, 2012.

[WSR19] Isaac Wang, Jesse Smith, and Jaime Ruiz. Exploring virtual agents for augmented reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[WSS20] Isabell Wohlgenannt, Alexander Simons, and Stefan Stieglitz. Virtual reality. *Business & Information Systems Engineering*, 62:455–461, 2020.

[YEY18] Gürkan Yildirim, Mehmet Elban, and Serkan Yildirim. Analysis of use of virtual reality technologies in history education: A case study. *Asian Journal of Education and Training*, 4(2):62–69, 2018.