Computers & Graphics 83 (2019) 1–10

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on ICAT-EGVE 2018

Effects of virtual agent and object representation on experiencing exhibited artifacts

Susanne Schmidt[a,*], Gerd Bruder[b], Frank Steinicke[a]

[a] Universität Hamburg, Germany
[b] University of Central Florida, USA

ARTICLE INFO

Article history:
Received 1 March 2019
Revised 6 June 2019
Accepted 14 June 2019
Available online 18 June 2019

Keywords:
Virtual agents
Virtual museum
Virtual/mixed reality
User evaluation

ABSTRACT

With the emergence of speech-controlled virtual agents (VAs) in consumer devices such as Amazon's Echo or Apple's HomePod, we have seen a large public interest in related technologies. While most of the current interactive conversational VAs appear in the form of voice-only assistants, other representations showing, for example, a contextually related or generic humanoid body are possible. In our previous work, we analyzed the effectiveness of different forms of VAs in the context of a virtual reality (VR) exhibition space. We found positive evidence that agent embodiment induces a higher sense of spatial and social presence. The results also suggest that both embodied and thematically related audio-visual representations of VAs positively affect the overall user experience. We extend this work by further analyzing the effects of the physicality of the agent's environment (i.e., virtual vs. real). The results of the follow-up study indicate some benefits of virtual environments, e.g., regarding user engagement and learning of visual facts. We also evaluate some interaction effects between the representations of the virtual agent and its surrounding and discuss implications on the design of exhibition spaces.

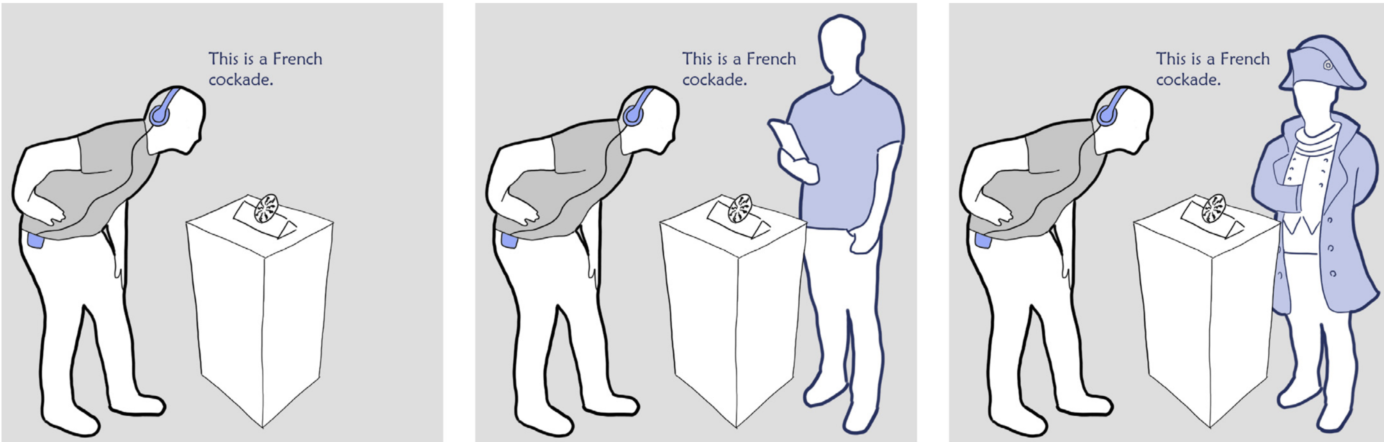
1. Introduction

Virtual agents (VAs) have come a long way from science fiction media to our daily social life. Driven by technical advances in the research fields of machine learning, internet of things, and virtual reality (VR), devices with embedded voice-controlled agents entered the consumer market to allow customers an intuitive and natural form of interaction with their smart home environments and as a means to access information from the internet [1]. Beyond home uses, smart services provided by VAs are popular as they can be accessed through ubiquitous smartphone technologies and can be implemented for professional applications such as in the form of educational audio guides in museums or audio-visual presentations for mixed media installations or exhibits. In particular, in situations where the demands for individual support or care exceed the supply of specialized trained personnel, such as museum guides, caregivers, or private assistants, these VAs are a promising solution that can complement human professionals [2]. Based on rapid development of smart services in recent years, it seems reasonable to assume that people will be confronted with an increasing amount of such services, which poses new challenges to the interface designers, particularly in terms of social interaction and integration.

While most of the currently used services are limited to audio or flat 2D visual representations, VR and Augmented Reality (AR) technology can add a new dimension by providing a 3D virtual body to complement the voice. Human-like VR/AR representations can enrich the communicative channels that convey the agent's status and intentions to interlocutors with gestures and other forms of social behaviors. Moreover, they can be registered spatially with their environment, which enables a more direct form of spatial interaction compared to voice-only interaction. This is particularly interesting in situations that have a strong spatial component such as art installations and museum exhibitions, since spatial relations are usually harder to communicate via speech than with gestures [3]. Therefore, it may be beneficial to provide a VA with a virtual body, which could also increase the user's feeling of co-presence, i. e., raising the visitor's sense of being together with the content on display. For museum exhibits this could be strengthened, for instance, by choosing a historical person as the agent's representation, as exemplified in Fig. 1. Through the encounter with a contemporary witness, visitors get to know the subject matter from a personal perspective, which may increase interest in the historical events as well as empathy with the people involved.

* Corresponding author.
E-mail address: schmidt@informatik.uni-hamburg.de (S. Schmidt).

https://doi.org/10.1016/j.cag.2019.06.002
0097-8493/



(a) **(b)** **(c)**

Fig. 1. Example museum application with (a) a traditional audio guide, (b) a generic embodied virtual guide, and (c) a content-related embodied virtual guide.

In our previous work, we presented a human-subject study that we performed in a historical exhibition context to understand the importance of different representations of virtual agents [4]. In this context, we analyzed the effectiveness of virtual museum guides with varying embodiment (embodied vs. disembodied) and thematic closeness (astronaut vs. museum guide) in the scope of a simulated exhibition related to the Apollo 11 mission. In particular, we were interested in the effects on the elicited sense of social presence, knowledge transfer, and the ability to communicate a sense of social competence and trust.

We extend this work by further analyzing the effects of the representation (i.e., virtual vs. physical) of the exhibit in focus. By including this additional factor, we aim to increase ecological validity of the results, since most traditional museums place real exhibits on display rather than relying on purely virtual visualizations.

Throughout the article we evaluate the following three research questions:

1. Do embodied virtual guides perform significantly better than voice-only guides in terms of co-presence, social presence, credibility, and the ability to impart knowledge?
2. Do thematically close content-related guides perform better than generic guides in terms of the above-mentioned metrics?
3. Is the performance of virtual guides affected by the physicality of surrounding objects?

The results of the user studies indicate benefits of embodied as well as thematically close audio-visual representations of virtual guides, both in the presence of virtual and physical exhibits. Higher scores in terms of user engagement and knowledge transfer also suggest advantages of including a virtual component in educational applications, either in the form of an embodied agent or as a virtual exhibit. We discuss implications and suggestions for user interface and content developers to design believable virtual agents in the context of both virtual and physical installations.

The remainder of this article is structured as follows. Section 2 gives an overview of related work on VAs in VR/AR as well as on embodiment and presence. In Section 3 we summarize the user study that we conducted in [4] and discuss the results. A follow-up study focusing on real exhibition spaces is detailed in Section 4. Section 5 concludes the article and discusses future research.

2. Related work

In this section, we cover work related to agent embodiment as well as presence and confidence in virtual agents.

Virtual Agents in VR/AR. Different forms of VAs were proposed and evaluated throughout Milgram's reality-virtuality continuum, which were surveyed by Holz et al. [5,6] and Norouzi et al. [2], showing the potential of VR/AR agents, but also challenges related to creating a high sense of social interaction and connection between users and virtual agents. For instance, Obaid et al. [7,8] showed that the physiological arousal of users in VR/AR depends on an agent's behavior associated with cultural differences, e. g., related to gaze behavior and interpersonal distances. Lee et al. [9] found that the proxemics during interaction with VAs in AR differs significantly from those between real humans, with users giving virtual agents more space than they would a real person. Kim et al. [10,11] showed that visual conflicts in AR such as occlusion and dual occupancy between virtual agents and physical objects can significantly impair their social connection with users. However, despite the challenges related to realistic and/or effective social interaction, a large number of applications could benefit from VAs [1,2]. For further information on VAs, we refer to Magnenat-Thalmann et al. [12], which provides a literature review of promising application fields for VAs including interactive virtual guides in cultural heritage sites, museums, art installations, and related fields.

Embodiment. A large body of literature focused on the question if and how virtual agents should be embodied for effective interaction with real humans. Dehn and van Mulken [13] presented a literature review on this subject, showing that the early prototypes of embodied agents in the last millennium had mixed effects on human-agent interaction, which sometimes would improve agent-based user interfaces while often they would not provide any benefits. A newer literature meta review by Yee et al. [14] showed benefits for virtual agents with a face over those implemented just by voice or via text on a computer screen. Moreover, they found that showing a face in general is more important than the realism of the visual presentation or the behavior of the agent. Even an abstract face of a virtual agent can provide important social cues for human-agent interaction, such as communicating visual attention due to the gaze direction of the eyes. Over the last decade, a large number of studies were conducted using VR technologies, which documented the psychological benefits of agent embodiment based on holographic or stereoscopic 3D displays, with recent work showing benefits for rapport as well as realistic social interaction [15,16]. Demeur et al. [17] showed that virtual agents in social situations can appear more believable and elicit a higher sense of competence and warmth if they are embodied. Recent studies by Kim et al. [10,11,16] indicate that it is more challenging for embodied virtual agents in AR to appear realistic and plausible to users. This is due to the fact that they are more affected by differences in visual appearance between their body and real-world objects as well as the fact that they generally have less control over their environment than is the case in VR, which can limit how virtual agents in AR are perceived and utilized.

Social and co-presence. A generalizable metric for the effectiveness of VAs in VR/AR is their ability to convey an illusion of being perceived as a real social entity sharing the same space with a real person, called *social presence* and *co-presence*. Co-presence denotes the sense of "being together" and social presence the sense of "being socially connected" [18]. Blascovich et al. define social presence as "the degree to which one believes that he or she is in the presence of, and dynamically interacting with, other veritable human beings" [19,20]. In general, the sense of "being there" is denoted *presence*, which can be further refined with the concepts of *place illusion* and *plausibility illusion* introduced by Slater [21]. Plausibility illusion indicates that "the scenario being depicted is actually occurring" with a "credible scenario and plausible interactions between the participant and objects and virtual characters in the environment." Various studies were conducted on VAs in VR/AR aimed at identifying effects of VA characteristics on the sense of social and co-presence during interaction using measures such as questionnaires, physiological responses, and behavioral differences, e.g., related to proxemics [22]. For instance, Lee et al. [9] found that more realistic multimodal feedback related to footstep vibrations of virtual agents that are transmitted through the floor using subwoofer devices could significantly improve subjective ratings of social and co-presence in AR. Moreover, they found that the limited field of view of current-state optical see-through AR head-mounted displays (HMDs) can negatively impact social and co-presence and cause less natural proxemic behavior near a virtual agent. Nowak and Biocca [23] evaluated different types of VAs and found to their surprise that a higher anthropomorphism of VAs reduced the sense of social and co-presence, which they explained stating that higher anthropomorphism might reinforce a person's expectations about realistic behaviors of the virtual agent, which were not met in their study. Chuah et al. [24] proposed hybrid VAs with partially physical body parts (legs) for medical applications, suggesting that a higher physicality of VAs could encourage higher social presence. This result is similar to what Kim et al. [16] found for robotic VAs. Kim et al. [25] further observed that the sense of social and co-presence depends on the personality of the real interlocutors, showing that extroverted participants reached a higher social presence with VAs than introverted participants.

3. User study with virtual exhibits

In this section, we summarize the user study that we conducted to understand the effects of a virtual agent's embodiment relative to the thematic context on the example of a virtual exhibition. Further details can be found in [4]. In the context of this study, we explored an exhibition, which simulated four episodes of the first manned moon landing. Each episode was presented by a different virtual guide in randomized order: (i) a generic virtual character or (ii) a thematically close content-related astronaut, each presented either as (iii) a disembodied voice (as known from voice-controlled agents such as Amazon's Echo) or (iv) a stereoscopic 3D embodied representation.

3.1. Participants

In total, 24 participants (17 male and 7 female; ages from 19 to 39, $M = 25.1$) participated in our experiment. All of them were students or staff members of the local Department of Engineering and Computer Science. None of the participants reported any visual or motor impairments that could affect the results of our experiment.

3.2. Material

The experiment was conducted in a CAVE environment with a scope of around 13 m^2 and three walls as well as the floor as projection surfaces. Four projectors were used to project rendered virtual imagery, each providing a resolution of 1920×1080 at a refresh rate of 120 Hz. In order to experience stereoscopic content, participants wore shutter glasses that were equipped with passive markers to track the user's position and orientation within the environment. The voice of the virtual guides was presented to participants via wireless noise-canceling headphones. Hence, participants were not restricted in their movement and were able to walk around virtual objects in the CAVE freely. Fig. 2(a) shows the experimental setup.

In our case study, we presented four episodes of the Apollo 11 mission, for which we used different models of a scaled-down Saturn V rocket with a launch pad, the interior of the Columbia command module, a scale model of the lunar module, and the moon surface with the American flag as well as scientific experiments (see Fig. 2(c)–(f)). If available, original footage such as a 3D scan of the command module and NASA photographs of the lunar surface was used in order to build detailed models.

We created four versions of the VA used in the experiment (see Fig. 2(b)):

1. The *embodied thematically close* character was modeled as an astronaut with a space suit. The astronaut's face was generated using original footage of Neil Armstrong.
2. The *embodied more generic* virtual character was designed to match a museum guide wearing a shirt and dress pants. In order to prevent any preference towards one of the guides due to sympathy, we used similar basic facial characteristics for the civilian guide. However, variations of the textures, facial hair and general hair style were made to ensure that the civilian and the astronaut were not perceived as the same person.
3. The *disembodied voice of the thematically close* astronaut character was identical to that condition except for the visual feedback of the agent.
4. The *disembodied voice of the more generic* character matched the embodied condition except for the visual feedback.

To increase the level of realism, we added idle behaviors to the embodied virtual guides, and they made eye contact with the user as a real guide would do in a one-on-one conversation.

In the embodied conditions, the agent's lip movements were matched with the spoken text via the Oculus Lip Sync plug-in. The audio track of the guides was created with the Oddcast Vocalware text-to-speech engine. For the astronaut, additional post-processing in Audacity was applied to simulate the sound of radio transmissions at that time.

The four episodes of the Apollo 11 mission provided educational information to the participants, narrated by the virtual guides. The assignment of a virtual guide to the four episodes was randomized. The educational content differed between the four episodes, but it was the same for all guides, except for the narrative point of view: The thematically close astronaut told the story from a first-person perspective and called "his" companions by their given names, while the more generic museum guide told the story from a third-person perspective. A detailed description of the four episodes can be found in [4].

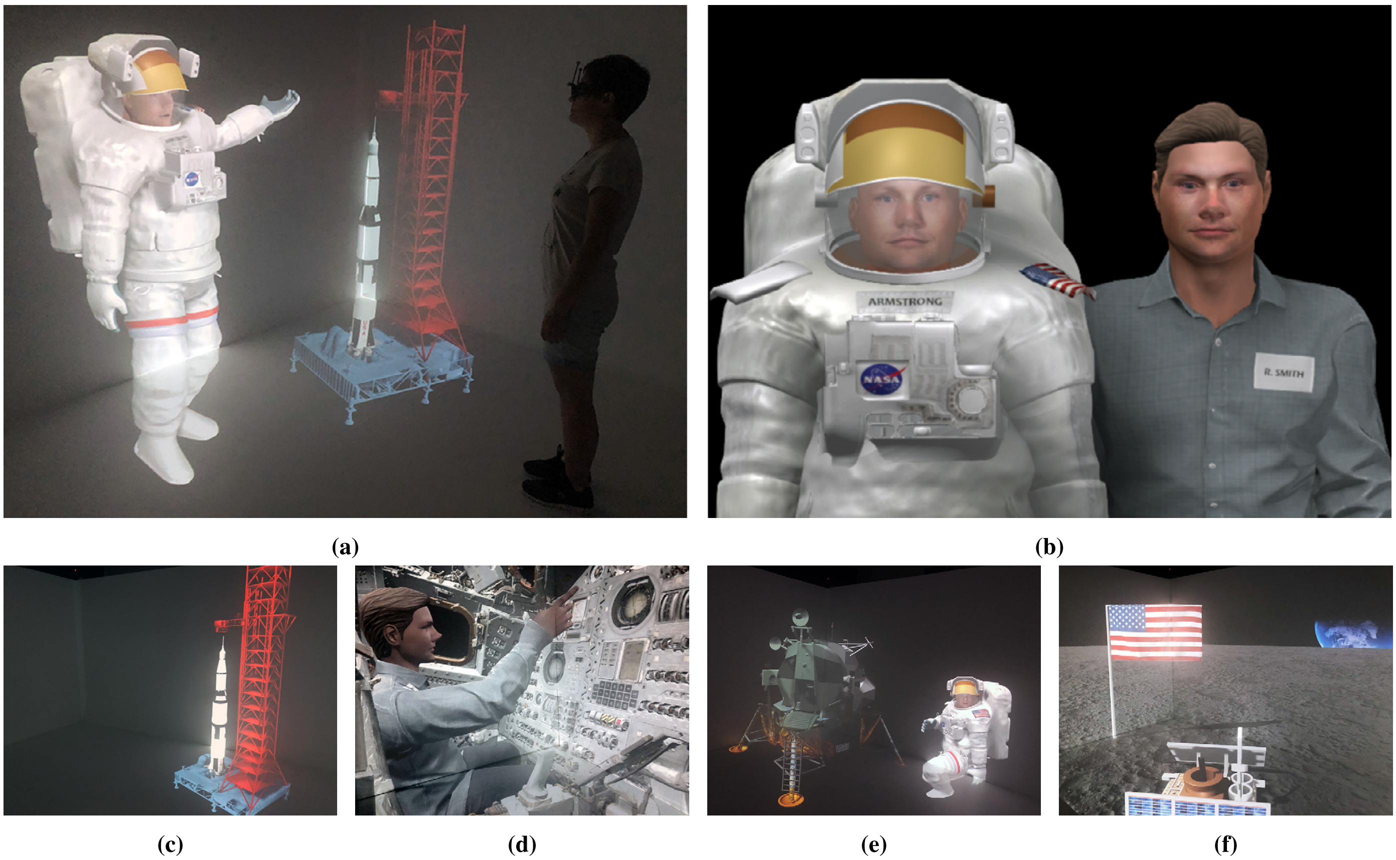


Fig. 2. (a) Photo showing the experimental setup, (b) the two guides in their embodied version, and (c)–(f) photos of the four episodes with exemplary guides.

3.3. Methods

For the first study, we used a within-subjects design based on two factors with two levels each: *agent embodiment* (embodied vs. disembodied) and *thematic closeness* (astronaut vs. museum guide). Each participant experienced all four episodes and all four agents described above in randomized order.

Prior to the study, each participant completed a consent form and a demographics questionnaire. Afterwards, participants were guided into the CAVE-like environment by following a virtual 3D floating globe. Participants were introduced to the display technology, had time to familiarize themselves with the system and the stereoscopic display, and then they were informed about the context of the study and the Apollo 11 mission scenario.

After this introductory phase, the main study started with the first of the four episodes of the Apollo 11 mission. Each episode took around three minutes to complete. Participants were allowed to move about the space in the experimental room freely. During the episodes, one of the four guides was present and gave a presentation on the virtual space models on exhibition in the CAVE.

After each episode, the participants were asked to rate their experience using subscales of the Temple Presence Inventory [26] as well as questionnaires that address the agent's credibility and the subjective knowledge gain.

We further ran participants through an "exam" on the presented educational content of the episode they just experienced, assessing how much of the information they actively perceived and could remember. The exam was chosen as a meaningful measure of the guides' quality, since museums usually have an educational mandate. While the visitor is not expected to learn all facts that are presented within an exhibition, the ability to provide interesting information that stick in the visitors' minds is of great value to any public educational institution. In this sense, the exam should give an idea on how successful a guide was to tell a memorable story rather than providing a generalizable percentage of learned facts. Initially, we planned for the exam to be completed without prior notice of the participants at the end of the study. However, a pre-study with ten participants revealed that only a minority of the users paid attention to any of the spoken text and the majority understood it more as an educational entertainment experience. We therefore decided to announce the exam before the study. For each episode a set of 12 questions was prepared, which were similar in terms of their memorizability. They were grouped into four categories: numerical, spatial, social, and visual facts. Numerical questions included sizes, weights, quantities, and periods of time. In spatial tasks, participants had to point at a specific location within a picture of the according scene. This location was described during the episode and was usually supported by a gesture in the embodied conditions. Social facts referred to stories that were experienced by the crew and members of the mission. Visual features were not mentioned by the guide, but could be observed in the presented scene. The exam was conducted orally to ensure that responses, which were guessed or already known before the study, could be identified.

After the exam was finished, participants were guided to the next episode and all steps were repeated. The second episode differed from the other scenes since participants were seated in the center of the CAVE. At the end of all episodes, participants were confronted with all four guides for a second time and had to compare them in an additional questionnaire. The entire study took around 45 to 60 minutes per participant.

Table 1
Main and interaction effects of the two factors *agent embodiment* and *thematic closeness* on each dependent variable. Asterisks indicate a statistically significant effect (* significant at .05 level, ** significant at .01 level or lower, *** significant at .001 level or lower).

	Presence				Learning					User Experience					
	Spatial	Active social	Social actor	Engagement	Numerical	Spatial	Social	Visual	Credibility	Attractiveness	Perspicuity	Efficiency	Dependability	Stimulation	Novelty
Agent embodiment	***	**	***	*	–	–	–	**	*	**	**	–	–	***	***
Thematic closeness	–	–	*	–	–	–	–	–	–	*	–	–	–	**	**
Embodiment * closeness	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

3.4. Results

We evaluated the effect of the two factors *agent embodiment* and *thematic closeness* on several subjective and objective measures using multiple two-way repeated measures ANOVAs. The normality assumption was not met in a few cases, however, the ANOVA tolerates moderate deviations from normality, as was shown in several studies [27–29]. A summary of all main and interaction effects can be found in Table 1.

3.4.1. Presence

Different aspects of presence were measured using the Temple Presence Inventory (TPI) [26]. We focused on four dimensions of the TPI: spatial presence, active social presence, presence as social actor, and presence as engagement. Each dimension involved three to seven items that were measured on a 7-point Likert scale. We ran a two-way repeated measures ANOVA that revealed a significant main effect of agent embodiment on spatial presence ($F(1,23) = 25.822, p < 0.001, \eta_p^2 = 0.529$), active social presence ($F(1,23) = 12.181, p = 0.002, \eta_p^2 = 0.346$), presence as social actor ($F(1,23) = 299.404, p < 0.001, \eta_p^2 = 0.929$), and presence as engagement ($F(1,23) = 7.516, p = 0.012, \eta_p^2 = 0.246$). Thematic closeness only showed one significant main effect on presence as social actor ($F(1,23) = 4.420, p = 0.047, \eta_p^2 = 0.161$). No other main effect or interaction effect was significant. The results of the TPI are illustrated in Fig. 3(a).

3.4.2. Learning

Scores of the oral exam were added up per participant and category, with a score of 3 corresponding to the maximum value of 100%. The results of one participant had to be removed from the data, because he admitted to know several of the tested facts even without the guides due to prior knowledge on the moon landing. The remaining scores were pooled according to the four categories as illustrated in Fig. 3(b). An ANOVA revealed a significant main effect of agent embodiment on the test scores in the category of visual facts ($F(1,22) = 8.933, p = 0.007, \eta_p^2 = 0.289$). Apart from this, no other effects on the learning results could be found.

In addition to the objective exam, we also wanted to learn more about the subjective impression of the participants regarding their knowledge gain through the guided presentations. After each episode, before the oral exam, we asked them to make a rough estimate on how many facts they are still able to recall now and in one week. We ran another ANOVA and found a significant main effect of embodiment on the perceived number of long-term memorized facts ($F(1,23) = 16.403, p < 0.001, \eta_p^2 = 0.416$), but not on the number of short-term memorized facts ($F(1,23) = 3.185, p = 0.088, \eta_p^2 = 0.122$).

3.4.3. Credibility

For evaluation of the credibility of guides, we used a scale introduced by McGloin et al. [30]. For each participant, an overall score was constructed from five questionnaire responses to the following bipolar adjective items: “unintelligent to intelligent”, “uninformed to informed”, “unreliable to reliable”, “incompetent to competent”, and “untrustworthy to trustworthy”. We analyzed the results with a two-way repeated measures ANOVA. Analysis revealed a main effect of agent embodiment on credibility ($F(1,23) = 5.842, p = 0.024, \eta_p^2 = 0.203$), indicating a significant difference between embodied guides ($M = 5.550, SD = 0.888$), and guides with voice only ($M = 5.254, SD = 1.030$).

3.4.4. User experience

Besides the aforementioned influence of agent embodiment and thematic closeness on perceived presence, agent credibility and learning, we were also interested in the general experience of users while interacting with the guides. For this purpose, we measured six dimensions of user experience with the UEQ [31]. Participants of the study were asked to provide ratings on 26 items using a 7-point Likert scale. We stressed the point that all responses should be based on the impression of the guide only, without including the virtual scene. This is because the virtual objects were only used in the context of the first study and are no inherent part of applications with AI agents in general. For example, a museum could also incorporate a virtual guide to present real physical exhibits instead of virtual ones; a scenario that was investigated in the follow-up study. We ran two-way repeated measures ANOVAs for the six dimensions of UEQ. We found a significant main effect of agent embodiment on attractiveness ($F(1,23) = 8.837, p = 0.007, \eta_p^2 = 0.278$), perspicuity ($F(1,23) = 8.307, p = 0.008, \eta_p^2 = 0.265$), stimulation ($F(1,23) = 26.527, p < 0.001, \eta_p^2 = 0.536$), and novelty ($F(1,23) = 82.786, p < 0.001, \eta_p^2 = 0.783$). Thematic closeness also showed a main effect on attractiveness ($F(1,23) = 7.212, p = 0.013, \eta_p^2 = 0.239$), stimulation ($F(1,23) = 10.291, p = 0.004, \eta_p^2 = 0.309$), and novelty ($F(1,23) = 10.505, p = 0.004, \eta_p^2 = 0.314$). No significant interaction effects between agent embodiment and thematic closeness were found. The results are illustrated in Fig. 3(d).

After a participant experienced all conditions, he was asked for a subjective ranking of the four different guides. Embodied guides were preferred by most of the participants, with 6 votes for the generic museum guide and 14 votes for the astronaut. In comparison, the unembodied generic guide took last place for 12 and the unembodied astronaut for 9 of the participants.

In a pre-study, a participant pointed out an unfair inequality between guides, because he perceived the condition with an embodied astronaut to be the only one dubbed by a real person, while the others were assumed to be generated by a text-to-speech engine. Since even the astronaut guides with and without body were rated differently, although the same artificially generated voice was used for both of them, we decided to pursue investigations on this aspect in the main study. For each guide participants had to decide whether the spoken text seemed to be produced by a text-to-speech engine or by a real speaker. For the unembodied astronaut, 45.8% of the participants assumed that the agent was synchronized by a real person. For the embodied astronaut, this was the case for even 62.5% of all participants. In contrast, the option of a real speaker was chosen by 37.5% of the participants for the embodied generic guide, and only by 33.3% for the unembodied generic guide.

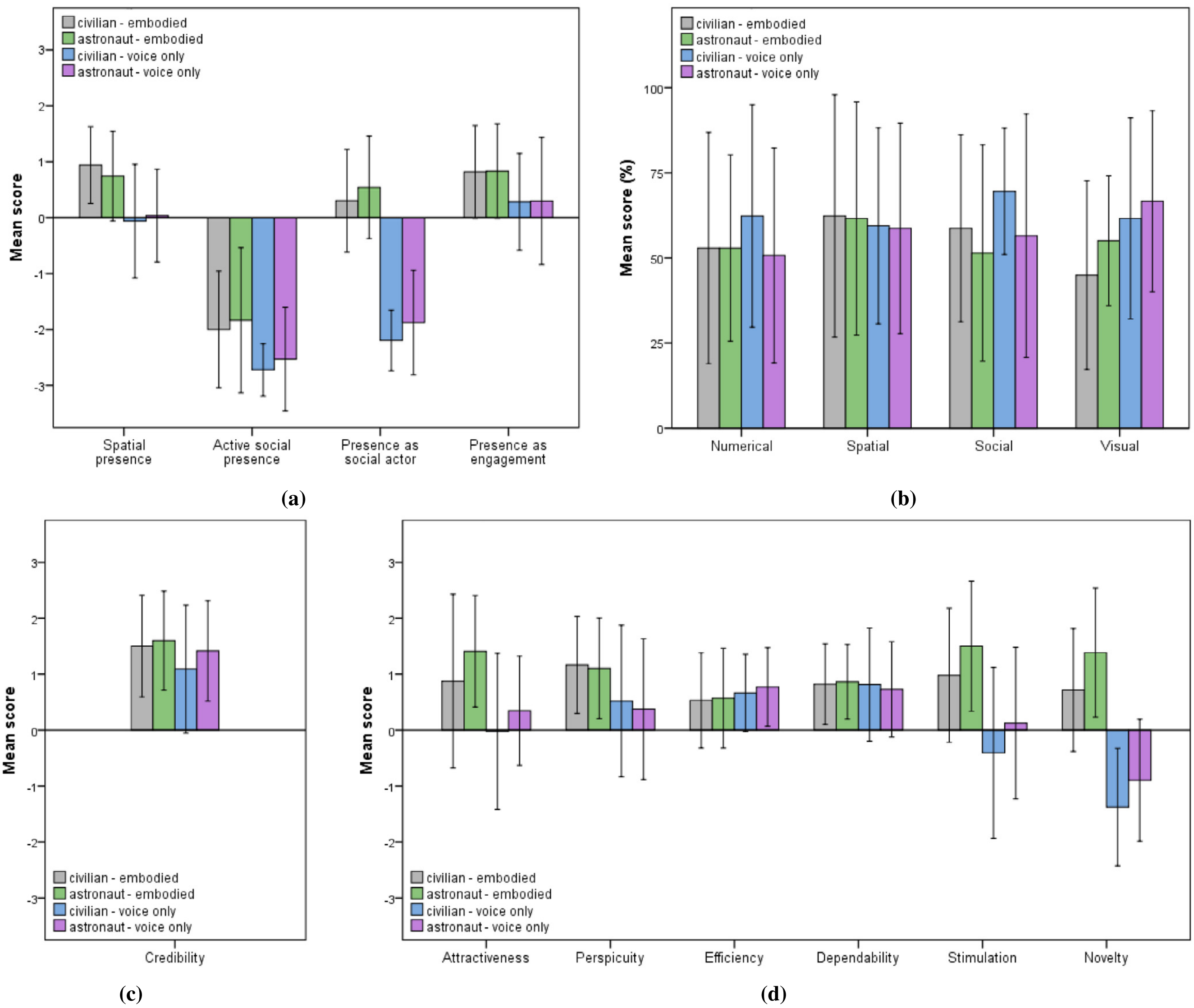


Fig. 3. Pooled results of (a) different presence measures, (b) learning results in four categories, (c) agent credibility, and (d) six dimensions of user experience. The vertical bars show the standard deviation.

3.5. Discussion

Even though our exemplary museum application did not include any forms of active interaction between the participants and the guide, the agent's embodiment had a positive effect on all measured presence dimensions. Through the presence of a second individual within the CAVE, participants felt significantly more spatially involved in the virtual environment. Participants also reported that the embodied guides caused more emotional responses such as laughing or smiling. In general, there was only little active interpersonal communication between users and guides in all conditions, however, this could be regulated by the introduction of additional interaction mechanisms such as voice commands. Whether this is desirable strongly depends on the application itself. In public settings such as a museum, speaking with a virtual agent may make users feel uncomfortable. In contrast, speaking with a personal assistant at home is already the custom and generally accepted. The most remarkable difference between embodied and unembodied guides can be observed in the scores of presence as social actor, sometimes also referred to as parasocial interaction. This measure of presence contains items that are related to crossing the border between the actual physical environment and the mediated environment in order to interact with the agent in real time [32]. Higher scores for embodied guides indicate that participants felt that their presence was noted by the agent and that he was establishing a connection to them. Although no complex reactions of the agent to the user's behavior were implemented, a feature as simple as making eye contact seems to be an effective method to create a sense of responsiveness and intimacy. Not only the agent's embodiment but also his thematic closeness had a main effect on presence as social factor. This positive effect could be caused by the first-person perspective of the astronaut, since the guide was not only imparting knowledge but was inviting the user to take part in his personal story.

Regarding credibility, all guides got mean scores in the upper range of the 7-point scale. Besides the realism of guides this could also be attributed to the fact that users do not expect museum guides to lie to them about the chronological order of historical events. Nevertheless, we found a significant effect of agent embodiment on the perceived credibility, indicating that embodied guides seemed to be even more competent and trustworthy.

Despite the exam was announced beforehand to the participants of the study, we expected different learning results for the four types of agents, in particular with regard to the different categories of information. However, this hypothesis could be confirmed only to some extent. Visual details such as the color of specific objects could be remembered better in conditions with a voice-only guide than in scenes with an embodied guide. We expected this outcome, since users tend to follow the agent's lip movements in the embodied condition and therefore could be more distracted from the actual scene. On the other hand, we hypothesized a positive effect of embodiment on the memorization of spatial information, however, such an effect could not be found in the data. In contrast to the results of the objective oral exam, participants subjectively perceived their gain of knowledge to be higher in the conditions with embodied guides, in particular in the long term. Indeed, the involvement of multiple modalities in the learning process as well as an increased presence in virtual environments were related to better learning results in previous studies [33]. A follow-up study that focuses on long-term effects of learning could resolve the question whether the subjective impression of participants can be sustained by an objective test.

We also expected the scores for social questions to be higher for guides with a personal connection to the stories. While no significant effect was found between the generic museum guide and the astronaut, Fig. 3(b) even indicates a trend in favor of the generic guide. The comment section of the questionnaires could give some indication of possible reasons for the observed behavior. Some participants stated that the astronaut was harder to understand due to the applied radio transmission effect.

Besides the problems in understanding the astronaut due to the added distortions, it was also mentioned that this effect made the astronaut sound more realistic than the generic guide. This impression was also confirmed by the responses to the question whether the audio was generated with a text-to speech engine or spoken by a real person. Besides the thematic closeness of the agent, his embodiment also affected the perceived realism of his voice in a positive manner. In spite of identical audio tracks, the presence of an embodied agent seems to distract the user from artifacts of speech synthesis and made the voice sound more natural. Therefore, the embodied astronaut was perceived to have a real voice by the majority of participants.

The perceived realism of the astronaut could also contribute to his positive reception by the participants of the study. In the usability questionnaires, the astronaut guides were rated as significantly more attractive, exciting and motivating, as well as innovative and creative. This is also true for embodied guides in comparison to guides with voice only. These results indicate that the extra effort that has to be made in order to implement a customized agent could be worthwhile to increase user satisfaction and improve the overall user experience.

4. Follow-up study with real exhibits

The experiment described in Section 3 provides insights into the effects of a virtual agent's embodiment in museum exhibitions with virtual exhibits. However, it remains open if a physical exhibit in combination with a virtual guide could further enrich the user experience. Hence, we conducted a follow-up study to replicate the scenario from the first experiment in an environment in which the virtual agent and real objects are blended into the same space.

4.1. Participants

For the follow-up study we recruited 24 participants (15 male and 9 female; ages from 20 to 46, $M = 26.5$), who did not participate in the first experiment. All participants were students or staff members of the local computer science department. Most of them had already some experience with VR/AR, since only two of them participated in a study involving VR or AR for the first time. As for the first experiment, we verified that participants do not suffer from any visual disorders that could interfere with the study procedure.

4.2. Material

In order to ensure comparability between both the first experiment and the follow-up study we used the identical technology setup at the same location as described in Section 3.2. Due to the different situation in this study, it was required to slightly adapt the scene. This particularly involved the presented exhibit, while the guides remained unchanged. For the follow-up study we decided to recreate the first episode using a plastic scale model of the Saturn V and its launcher as illustrated in Fig. 4(a). The physical rocket featured the same visual details as its virtual equivalent with half the overall size. Due to the smaller height of 77.5 cm it was placed on a white box and therefore could be examined by the participants of the study on eye level similar to a real exhibit in museums. As in the first experiment, the scale model was positioned in a corner of the CAVE, as illustrated in Fig. 2(a).

4.3. Methods

Due to the hardware constraints we focused on the first episode only, and hence, the follow-up study followed a between-subjects design with two independent variables: *agent embodiment* (embodied vs. disembodied) and *thematic closeness* (astronaut vs. museum guide). Participants were randomly assigned to one of the four resulting conditions.

The introduction was carried out as in the first experiment, including a consent form, a demographic questionnaire and the staging of the exhibition scenario. For participants who experienced a voice-only condition we omitted a demonstration of the stereoscopic display since none of the presented objects were virtual. All of the participants were instructed to imagine visiting a real space museum and to behave as naturally as possible. As in a real museum, participants were allowed to move freely within the exhibition space, but were prohibited from touching the exhibit. After the introduction, the selected guide appeared and presented the first episode of Apollo 11 as described before. As in the first study, the episode took around 3 minutes and was followed by a number of questionnaires that addressed different presence scales, the guide's credibility and the subjective knowledge gain. Afterwards, we performed the oral test using the same questions as in the first iteration. The experiment was concluded by some final questions regarding the user experience. As described above the procedure was slightly different from the first experiment since participants experienced only one of the guides before rating their experience.

4.4. Results

To evaluate the effect of presenting a real exhibit instead of a virtual one, we compared the observations of the first experiment's first episode with the follow-up study, which used the identical material and methods. Therefore, the data gathered in the first episode can be treated as obtained in a between-subjects design like in the follow-up study. Hence, we also considered the factor called *virtuality* with two levels *virtual exhibit* and *real exhibit*, and ended up with a $2 \times 2 \times 2$ design. The data was analyzed using a three-way ANOVA with the three factors *agent embodiment*, *thematic closeness* and *virtuality*. An overview of all main and interaction effects is presented in Table 2.

4.4.1. Presence

For evaluating the presence we excluded the subscale of spatial presence since half of the participants in the second experiment did not experience virtual content at all and were

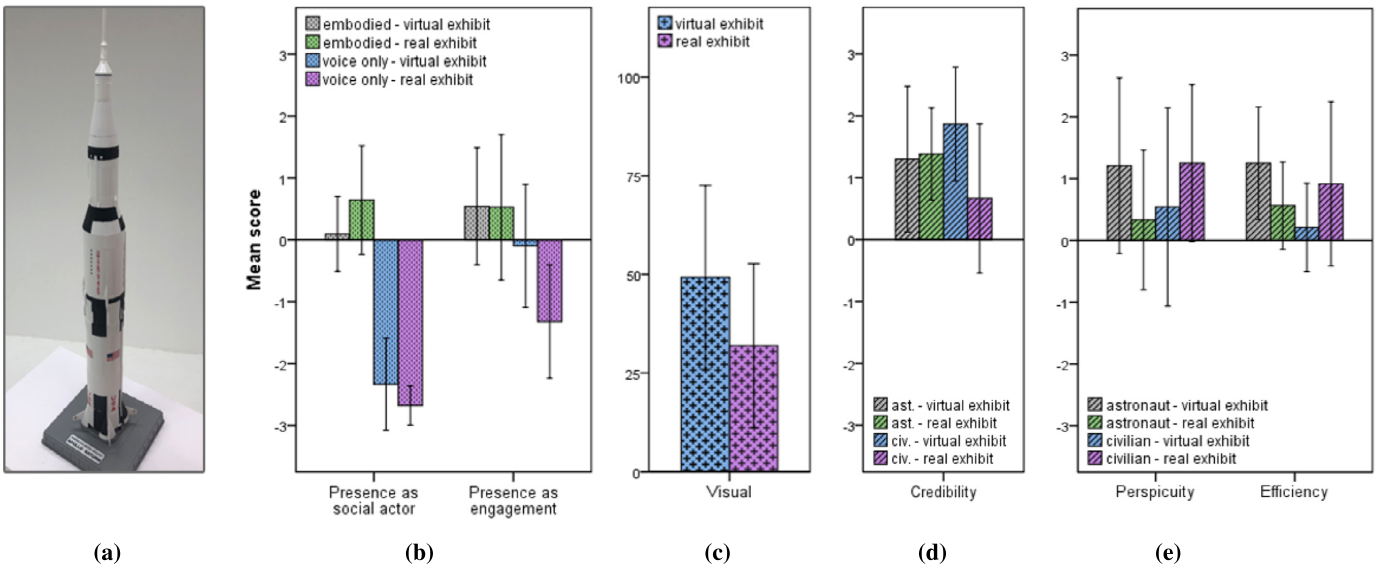


Fig. 4. Pooled results of the merged experiments including (a) different presence measures, (b) learning results in the visual category, (c) agent credibility, and (d) two dimensions of user experience. The vertical bars show the standard deviation.

Table 2
Pooled results of the second experiment and the first episode of the initial experiment. Asterisks indicate a statistically significant main or interaction effect of the three factors *agent embodiment*, *thematic closeness*, and *exhibit virtuality* on the corresponding dependent variable (* significant at .05 level, ** significant at .01 level or lower, *** significant at .001 level or lower).

	Presence			Learning					User Experience					
	Active social	Social actor	Engagement	Numerical	Spatial	Social	Visual	Credibility	Attractiveness	Perspicuity	Efficiency	Dependability	Stimulation	Novelty
Agent embodiment	***	***	***	–	–	–	–	–	***	**	–	–	***	***
Thematic closeness	–	–	–	–	–	–	–	–	*	–	–	–	*	*
Exhibit virtuality	–	–	*	–	–	–	**	–	–	–	–	–	–	–
Embodiment * closeness	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Closeness * virtuality	–	–	–	–	–	–	–	*	–	*	*	–	–	–
Embodiment * virtuality	–	*	*	–	–	–	–	–	–	–	–	–	–	–
Embodiment * closeness * virtuality	–	–	–	–	–	–	–	–	–	–	–	–	–	–

therefore not able to make valid statements on this dimension of presence. We found significant main effects of embodiment on all remaining subscales, namely active social presence ($F(1,40) = 12.813$, $p = 0.001$, $\eta_p^2 = 0.243$), presence as social actor ($F(1,40) = 212.119$, $p < 0.001$, $\eta_p^2 = 0.841$), and presence as engagement ($F(1,40) = 17.982$, $p < 0.001$, $\eta_p^2 = 0.310$). In contrast to the first experiment, thematic closeness did not show any significant main effects.

Virtuality also showed a significant main effect on engagement ($F(1,40) = 4.446$, $p = 0.041$, $\eta_p^2 = 0.100$). Furthermore, the ANOVA revealed significant interaction effects between virtuality and embodiment both on presence as social actor ($F(1,40) = 5.115$, $p = 0.029$, $\eta_p^2 = 0.113$), and presence as engagement ($F(1,40) = 4.248$, $p = 0.046$, $\eta_p^2 = 0.096$). The results involving virtuality are illustrated in Fig. 4(b).

4.4.2. Learning

The results of the oral exam were prepared for the analysis as described in Section 3.4.2. A three-way ANOVA resulted in a significant main effect of virtuality on test scores in the visual category ($F(1,39) = 7.574$, $p = 0.009$, $\eta_p^2 = 0.163$). No other significant effects on objective and subjective learning results could be found.

4.4.3. Credibility

Credibility scores, which were computed using the approach suggested by McGloin et al. [30], were also analyzed using an ANOVA. While no significant main effect could be found for any of the three factors, the ANOVA revealed a significant two-way interaction effect between virtuality and thematic closeness ($F(1,40) = 4.398$, $p = 0.042$, $\eta_p^2 = 0.099$). The interaction between both factors is visualized in Fig. 4(d).

4.4.4. User experience

As in the first experiment, we found significant main effects of agent embodiment on attractiveness ($F(1,40) = 20.169$, $p < 0.001$, $\eta_p^2 = 0.335$), perspicuity ($F(1,40) = 9.158$, $p = 0.004$, $\eta_p^2 = 0.186$), stimulation ($F(1,40) = 36.507$, $p < 0.001$, $\eta_p^2 = 0.477$), and novelty ($F(1,40) = 106.997$, $p < 0.001$, $\eta_p^2 = 0.728$). Thematic closeness also showed a main effect on attractiveness ($F(1,40) = 5.322$, $p = 0.026$, $\eta_p^2 = 0.117$), stimulation ($F(1,40) = 4.711$, $p = 0.036$, $\eta_p^2 = 0.105$), and novelty ($F(1,40) = 4.644$, $p = 0.037$, $\eta_p^2 = 0.104$). In addition, two significant interaction effects between virtuality and thematic closeness on perspicuity ($F(1,40) = 4.535$, $p = 0.039$, $\eta_p^2 = 0.102$) and efficiency ($F(1,40) = 6.058$, $p = 0.018$, $\eta_p^2 = 0.132$) could be found.

4.5. Discussion

We found significant differences for the virtuality of the exhibit as well as interactions between the virtuality and both embodiment and thematic closeness of the agent, as summarized in Fig. 4(b)–(e). In contrast to the first experiment, no significant main effects of thematic closeness on presence as social actor as well as embodiment on learning of visual facts and credibility were found in the aggregated data, which might be due to the reduced sample size.

One of the most interesting results from the study is the observed interaction effect between the agent's embodiment and the exhibit's virtuality on the subjective measure of presence as social actor, or parasocial interaction. As described above, this dimension of presence relates to a cross-over between the actual physical environment of the user and the mediated environment. Our analysis of the collected data of both experiments revealed that embodied guides achieved higher scores when displayed alongside a physical exhibit. Therefore, the physical exhibit may have supported the transfer of the virtual guide to the real environment of the participant. On the other hand, the parasocial presence was rated higher for the virtual exhibit than for the real one for conditions featuring a voice-only guide. This is interesting as it may indicate a reverse effect compared to the previously reported effect. As the audio guide was not embodied in the actual physical environment, a virtual exhibit may have helped the user feel more present in the virtual environment of the guide, therefore again bridging the gap between the user and the guide.

Another interaction effect between the embodiment of the agent and the virtuality of the exhibit was found for engagement. While the engagement ratings were similar for the embodied guides, they were significant lower for audio guides in conjunction with real exhibits. Participants assigned to this condition did not experience any virtual content, therefore being the closest to a traditional exhibition scenario. Though participants of the first experiment were explicitly asked to focus on the guide during their evaluation, the overall context with the virtual exhibit seemed to have an influence on their engagement as well.

A lack of engagement in the group of participants experiencing an audio guide with a real exhibit may also have contributed to the lower performance in the oral test with regard to visual facts. Overall, participants in the conditions with a virtual exhibit could remember more visual facts than participants in the conditions with a real exhibit. Though this effect could also be attributed to the differences in size and visual details of the real exhibit, the informal comments during the oral exam support another conclusion. In the second study, several participants who experienced an embodied guide reported that they were more interested in the virtual guide than the physical rocket, therefore not paying attention to visual features of the latter. Furthermore, in the first experiment, some participants who were assigned to a condition with an audio guide stated that they payed less attention to what was said since they preferred to explore the virtual rocket. It can be assumed that both reported behaviors eventually caused the effect which is shown in Fig. 4(c).

Another significant difference between the first and the second experiment was found for two of the six user experience scales. In the second study using a real exhibit, the generic museum guide was uprated while participants of the first experiment with virtual exhibits provided higher scores in favor of the astronaut. This interaction between thematic closeness and virtuality may provide an indication that a generic museum guide, which actually could be found in a real museum, fits in better with a real exhibition room than a content-related guide, whose presence is unusual for visitors of a museum. On the other hand, an environment with a virtual exhibit already is an exception to the norm and therefore, the presence of Neil Armstrong as a tour guide might be more relatable. However, this interpretation is limited in view of the fact that the results only apply to the scales of perspicuity and efficiency, while participants of both experiments preferred the astronaut in terms of stimulation and novelty.

5. Conclusion

In this article, we summarized two user studies, which investigated the effectiveness of different representations of virtual agents in an exhibition scenario. We analyzed the effects of two factors, embodiment (embodied vs. disembodied) and thematic closeness (astronaut vs. museum guide), on a number of variables that are relevant to the museum domain, including social presence, guide credibility, knowledge transfer, and visitor experience. In this context, we aimed to examine whether the costly and time-consuming implementation of embodied agents and their customization to a specific application give a competitive edge over common AI agents with audio only. The first study, which is presented in greater detail in [4], was conducted in a virtually simulated exhibition room addressing the Apollo 11 mission. In order to ensure ecological validity, we replicated the scene in a real exhibition space, and analyzed the effects in a second study.

In the pooled data of both experiments, we found significant differences between audio guides and embodied guides with regard to all presence measures as well as a subset of user experience scales, including perceived perspicuity, attractiveness, stimulation, and novelty. All effects were in favor of the embodied guide and therefore could justify the extra effort that is necessary to model and animate such an agent. This option should be taken into consideration for all applications, in which user experience is of top priority and the usage of additional technology such as a projector is reasonable, e.g., in public installations.

The content-specific guide in the form of an astronaut achieved higher scores in the dimensions of attractiveness, stimulation, and novelty. As a representative of historically relevant guides, we also expected an increase in both the credibility and the knowledge transfer, since visitors may emphasize with the guide's feelings and emotionally engage with him because of the personal connection to the told story. This hypothesis could not be confirmed based on the results of both user studies, however, credibility was rated slightly higher for the astronaut guide than for a generic museum guide. While no positive effects on learning could be found for the astronaut, visitors of a museum might be attracted by the more innovative guide representation and therefore pay more attention to the related exhibit. Furthermore, different results may be achieved in a field study within a real museum since according to the qualitative feedback many participants did not have a deep interest in the Apollo 11 mission and instead participated because of their interest in VR technology.

The virtuality of exhibits showed significant main effects on presence as engagement as well as the rate of remembered visual facts. The first result emphasizes general advantages of using virtual content in the museum context. Participants assigned to the condition with a real exhibit and a voice-only guide were significantly less engaged than participants of any other condition. This lack of engagement may also have contributed to the latter result, since participants who experienced a real exhibit apparently paid less attention to the visual details and therefore could remember only a few. Besides these differences between a virtual and a real exhibition space, most of the results of the first study could be reproduced within the second study, suggesting that the described positive effects of both embodied and content-related guides also apply to traditional museums with real exhibits.

Though virtual agents are emerging in various domains, we chose the environment of an exhibition in order to gain initial insight into the effectiveness of different agent representations. Some of the results may be applicable to other domains, too, as the considered aspects are relevant not only in the context of exhibitions. For example, high spatial and social presence values increase the VA's ability to be perceived as a real social entity, and therefore contribute to any social experience that involves VAs. The positive effects on variables such as attractiveness and stimulation are also of high value for other applications, as user experience is a key aspect for most human-computer interfaces. On the other hand, knowledge transfer is one of the more specific aspects in the presented studies, which may be less relevant to other domains. Instead, there might be additional application-specific factors to be included. For example, in health care for children a virtual expert such as a doctor could be compared to a less intimidating agent such as a mascot. In this domain, agent credibility and social presence are still of great importance, but other variables such as the release of fears should also be brought into focus. Additional studies are necessary to fully answer the questions which agents perform best in different scenarios.

Acknowledgments

The work presented in this publication is supported in part by the Office of Naval Research (ONR) Code 34 under Dr. Peter Squire, Program Officer (ONR award N00014-17-1-2927). The work is also supported in part by the National Science Foundation (NSF) under Grant Number 1800961 (Dr. Tonya Smith-Jackson).

References

[1] Kim K, Billinghurst M, Bruder G, Duh HB-L, Welch G. Revisiting trends in augmented reality research: a review of the 2nd decade of ISMAR (2008–2017). In: Proceedings of the IEEE TVCG; 2018. p. 1–16.
[2] Norouzi N, Kim K, Hochreiter J, Lee M, Daher S, Bruder G, et al. A systematic survey of 15 years of user studies published in the intelligent virtual agents conference. In: Proceedings of the IVA; 2018.
[3] Alibali MW. Gesture in spatial cognition: expressing, communicating, and thinking about spatial information. Spat Cogn Comput 2005;5(4):307–31.
[4] Schmidt S, Bruder G, Steinicke F. Effects of embodiment on generic and content-specific intelligent virtual agents as exhibition guides. In: Proceedings of the international conference on artificial reality and Telexistence and Eurographics symposium on virtual environments (ICAT-EGVE). The Eurographics Association; 2018.
[5] Holz T, Campbell AG, O'Hare GMP, Stafford JW, Martin A, Dragone M. MiRA – mixed reality agents. Int J Hum Comput Stud 2011;69(4):251–68.
[6] Holz T, Dragone M, O'Hare GMP. Where robots and virtual agents meet: a survey of social interaction research across Milgram's reality-virtuality continuum. Int J Soc Robot 2009;1(1):83–93.
[7] Obaid M, Niewiadomski R, Pelachaud C. Perception of spatial relations and of coexistence with virtual agents. In: Vilhjálmsson H, Kopp S, Marsella S, Thórisson K, editors. Proceedings of the international conference on intelligent virtual agents. Lecture notes in computer science, 6895. Springer Berlin Heidelberg; 2011. p. 363–9.
[8] Obaid M, Damian I, Kistler F, Endrass B, Wagner J, André E. Cultural behaviors of virtual agents in an augmented reality environment. In: Nakano Y, Neff M, Paiva A, Walker M, editors. Proceedings of the international conference on intelligent virtual agents. Lecture notes in computer science, 7502. Springer Berlin Heidelberg; 2012. p. 412–18.
[9] Lee M, Bruder G, Höllerer T, Welch G. Effects of unaugmented periphery and vibrotactile feedback on proxemics with virtual humans in AR. In: Proceedings of the IEEE TVCG, 24; 2018. p. 1525–34.
[10] Kim K, Maloney D, Bruder G, Bailenson JN, Welch GF. The effects of virtual human's spatial and behavioral coherence with physical objects on social presence in AR. Comput Anim Virt Worlds 2017;28(3–4):e1771.
[11] Kim K, Bruder G, Welch G. Exploring the effects of observed physicality conflicts on real–virtual human interaction in augmented reality. In: Proceedings of the ACM VRST; 2017. p. 1–7.
[12] Magnenat-Thalmann N, Papagiannakis G, Chaudhuri P. Applications of interactive virtual humans in mobile augmented reality. In: Furht B, editor. Encyclopedia of multimedia. Springer; 2008. p. 362–8.
[13] Dehn D, van Mulken S. The impact of animated interface agents: a review of empirical research. Int J Hum Comput Stud 2000;52:1–22.
[14] Yee N, Bailenson JN, Rickertsen K. A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In: Proceedings of the of ACM CHI; 2007. p. 1–10.
[15] Wang Y, Lucas G, Khooshabeh P, de Melo CM, Gratch J. Effects of emotional expressions on persuasion. Social Infl 2015;10(4):236–49.
[16] Kim K, Nagendran A, Bailenson JN, Raij A, Bruder G, Lee M, et al. A large-scale study of surrogate physicality and gesturing on human–surrogate interactions in a public space. Front Robot AI 2017;4(32):1–20.
[17] Demeure V, Niewiadomski R, Pelachaud C. How is believability of a virtual agent related to warmth, competence, personification, and embodiment? Presence Teleop Virt 2011;20(5):431–48.
[18] Harms C, Biocca F. Internal consistency and reliability of the networked minds measure of social presence. In: Proceedings of the annual international presence workshop; 2004. p. 246–51.
[19] Blascovich J. Social influence within immersive virtual environments. In: Schroeder R, editor. The social life of avatars. Springer London; 2002. p. 127–45. ISBN 978-1-85233-461-1. Computer Supported Cooperative Work.
[20] Blascovich J, Loomis J, Beall AC, Swinth KR, Hoyt CL, Bailenson JN. Immersive virtual environment technology as a methodological tool for social psychology. Psychol Inq 2002;13(2):103–24.
[21] Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philos Trans R Soc Lond B Biol Sci 2009;364(1535):3549–57.
[22] Fox J, Ahn SJ, Janssen JH, Yeykelis L, Segovia KY, Bailenson JN. Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. Hum Comput Inter 2014;30(5):401–32.
[23] Nowak K, Biocca F. the effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. Presence 2003;12(5):481–94.
[24] Chuah JH, Robb A, White C, Wendling A, Lampotang S, Kopper R, et al. Exploring agent physicality and social presence for medical team training. Presence 2013;22(2):141–70.
[25] Kim K, Bruder G, Maloney D, Welch G. The influence of real human personality on social presence with a virtual human in augmented reality. In: Proceedings of the ICAT-EGVE; 2016. p. 115–22.
[26] Lombard M, Ditton TB, Weinstein L. Measuring presence: the temple presence inventory. In: Proceedings of the PRESENCE 2009; 2009. p. 1–15.
[27] Glass GV, Peckham PD, Sanders JR. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. Rev Educ Res 1972;42(3):237–88.
[28] Harwell MR, Rubinstein EN, Hayes WS, Olds CC. Summarizing Monte Carlo results in methodological research: the one- and two-factor fixed effects ANOVA cases. J Educ Stat 1992;17(4):315–39.
[29] Lix LM, Keselman JC, Keselman H. Consequences of assumption violations revisited: a quantitative review of alternatives to the one-way analysis of variance F test. Rev Educ Res 1996;66(4):579–619.
[30] McGloin R, Nowak KL, Watt J. Avatars and expectations: influencing perceptions of trustworthiness in an online consumer setting. PsychNol J 2014;12.
[31] Laugwitz B, Held T, Schrepp M. Construction and evaluation of a user experience questionnaire. In: Proceedings of the USAB. Springer; 2008. p. 63–76.
[32] Lombard M, Ditton TB, Crane D, Davis B, Gil-Egui G, Horvath K, et al. Measuring presence: a literature-based approach to the development of a standardized paper-and-pencil instrument. In: Proceedings of the PRESENCE 2000, 240; 2000. p. 2–4.
[33] Mikropoulos TA. Presence: a unique characteristic in educational virtual environments. Virt Real 2006;10(3–4):197–206.