Influence of Visual Appearance of Agents on Presence, Attractiveness, and Agency in Virtual Reality

Marius Butz[1(✉)], Daniel Hepperle[1,2], and Matthias Wölfel[1,2]

[1] Faculty of Computer Science and Business Information Systems, Karlsruhe University of Applied Sciences, Karlsruhe, Germany
{buma1060,daniel.hepperle,matthias.woelfel}@h-ka.de

[2] Faculty of Business, Economics and Social Sciences, Stuttgart, Germany

Abstract. The way a system in a virtual environment interacts with the user through an agent-mediated interface could potentially influence the overall perception of the system or certain aspects of it. To investigate which type of visualization influences presence, attractiveness, and sense of agency, we designed four different visualizations of agent-mediated interfaces, namely audio-only without embodiment, as an object, as an anthropomorphic object, and with a human appearance. Our results show that presence and agency are not affected by the type of visualization of the agent, while perceived attractiveness shows differences in hedonic as well as in pragmatic quality.

Keywords: Agent-mediated interfaces · Visual appearance of avatars

1 Introduction

The way a system in a virtual environment interacts with the user through an agent-mediated interface can reduce the need to apply specific design paradigms associated with more traditional human-computer interfaces. The value of conversational user interfaces is that they provide a form of interaction and exchange in an almost natural way that simulates human-to-human verbal communication. Also for object manipulation a speech interface holds several advantages for example in regard to ease of learning and uncomplicated handling [11]. While the representation of the conversational partner is limited by physical appearance and can vary only slightly by dress, the representation of the conversational partner in virtual space is free of such constraints. The virtual representation of the interlocutor can be changed in a variety of ways including no representation at all, human-like, or in the form of an animal or object, and can be customized depending on the situation, narrative, experimental setup (e.g., for human behavior studies), or user preferences. These variations offer the interlocutor the chance to fulfill a specific role, which, in addition to the use of language, behavior, etc., is also conveyed at least in part by its visual appearance.

M. Wölfel et al. (Eds.): ArtsIT 2021, LNICST 422, pp. 44–60, 2022.
https://doi.org/10.1007/978-3-030-95531-1_4

Adapting the visual appearance of the interlocutor to the respective situation and environment can support participative perceptions, even unconsciously, including

Familiarity: sense of familiarity between an agent and a user provides a shared context that creates a mental model for the user to know what to expect from the interlocutor [3]

Competency: the appearance convey a sense of expertise (e.g. being taught about space by an astronaut) [16]

Branding: the avatar can act as a brand ambassador like Mickey Mouse or the Michelin manikin [5,15]

Acceptance: can promote the suspension of disbelief so that situations that are different from a real situation have a higher level of acceptance; e.g. nobody wonders why Superman can fly while a teacher cannot [1]

Varying the visual properties of an agent offers potential for interaction concepts that would not be possible in the real world. For example, a wizard could make objects appear out of the blue, or an anthropomorphized lamp could highlight important aspects by illuminating them [25].

The interlocutor can be either *autonomous* or *controlled by a person*, in the first case it is usually called an *agent*, in the second case an *avatar*. Avatars are common in social VR applications such as Facebook's Horizon, VRChat, AltspaceVR, or Mozilla's Hubs. A comprehensive overview of the state of the art in the visual representation of avatars in social VR applications can be found in [10]. Agents are used in many applications; for example, they can provide information, guidance, or feedback in a virtual museum or classroom context.

2 Related Work

While there are several works related to avatar and agent visual appearance in 2D environments (predominantly on video games), there is little research related to VR environments. McDonnell et al. [14] compared ten different rendering styles, ranging from a toon pencil to an actual virtual human, as well as an auditory-only rendering on a 2D monitor. They found that participants performed better on lie detection in the audio-only case than when a character was rendered. In their paper, they point out that this might be related to the participants focusing mainly on the visual appearance of the character than on what was being said. They also found that cartoon-style characters were rated as more appealing than human-style characters.

In a more dated study, also on a 2D monitor, Gulz and Haake [7] state that individuals who prefer *task-oriented* communication do not prefer a particular visual style of the avatar, but individuals who prefer *relationship-oriented* communication prefer an iconic visualization style over a realistic one.

In the study by Forlizzi et al. [6], female-looking, male-looking, and abstract agents were compared with each other with the result that human-looking agents were rated higher than abstract agents. In addition, female-looking agents were preferred over male-looking agents. They also found that gender stereotypes play a role in the expectations of an agent, even in those cases when gender cues are minimal.

While these results can provide initial guidance regarding agent and avatar appearance for immersive VR, it should be noted that results on 2D screens do not necessarily translate to VR settings or effects such as the uncanny valley are more pronounced in head-mounted VR as opposed to screen settings [9].

Bergman et al. [4] compared two different agents, one robot-like and one human-like for two types of gesture usage (with and without) in a qualitative user study. Eighty participants took part in the study, divided into four groups (two types of agents and two types of gestures). All participants started with a short video of a self-introduction of the agent, then they evaluated the first impression using a questionnaire (first measurement). In the second phase, the agent described a building in six sentences, which was again evaluated using a questionnaire (second measurement). The results indicate that the perceived interpersonal warmth is higher for the robot-like agent in the first measurement than for the human-like one. However, after the second measurement, the perceived warmth of the robot-like agent is lower than before, while it remains constant for the human-like agent. Competence is perceived significantly higher for both agent types when the agent was able to gesticulate.

Lee et al. [13] investigated whether *user performance* depended on agent appearance (actual tutor vs. 3D annotation) on three different tasks (navigating through a maze, stretching exercises, and controlling a crane). User performance was measured in execution time and task precision. They found that the 3D annotators had a higher precision in the maze task and a lower execution time in the stretching exercises. Regarding user behavior, it was found that participants in the tutoring group attempted to mimic the behavior of the virtual tutor, while participants in the annotation group attempted to meet the conditions of success.

A study by Torre et al. [24] examined differences in reliance on different emotional expressions (smiling face, positive voice modulation, or both) for a cartoonish and a photorealistic agent. The evaluation was based on behavioral data within a survival task, assessments with a questionnaire, and qualitative comments. For this purpose, a hypothetical accident scenario was created in a desert and on the moon, where participants were asked to rank six functional objects according to their importance for survival. Subsequently, the virtual agent, originally intended to serve as a navigation assistant, suggested a different (inverted) order of the user rated items, and the participants were asked to create a final order dependent on the agents suggestions. The differences between the order of the participant(s) and the agent and how many item ratings were adopted by the agent were used as the basis for behavior-based trust. The result show increased trust in an agent with congruent, neutral expression, which—according to the authors—is due to the extreme situation (stranded in the desert or on the moon), while in the opposite case the expression were deemed as sarcastic or inappropriate.

While most studies about the representation of agents and avatars have been conducted on 2D monitors, little attention has been paid to more immersive technologies such as head-mounted VR or AR. As their representation and perception might differ in VR or AR in particular on its ability of non-verbal communication (gaze direction, facial expressions, or body language) it is important to investigate the impact of different forms of visual presentation for agents and avatars [21].

3 Hypotheses

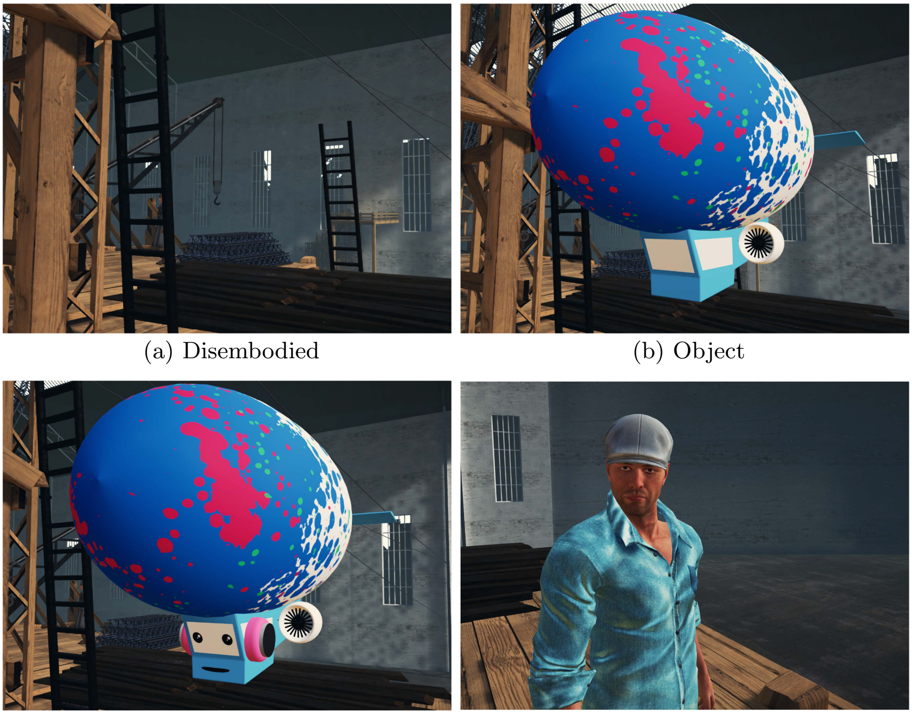
Since there has been little work on the visual representation of agents in immersive VR applications, we attempt to answer its influence on the three fundamental aspects presence, attractiveness, and sense of agency, as formulated in the following hypotheses:

H1 Presence is influenced by the type of visual appearance of the agent.
H2 Attractiveness is influenced by the type of visual appearance of the agent.
H3 Sense of agency is influenced by the type of visual appearance of the agent.

4 VR-Experience

To evaluate the hypotheses, a virtual environment is needed that provides a good match between the environment and the visual representation of the agent. A good match is given when the style and the content correspond. Our goal was to cover different interaction types and to include situations in which the agent provides information (passive user) and those in which it instructs what to do (active user). Due to the increased complexity and reduced resemblance of the test situation between participants, we decided against evaluating a two-way conversation between participant and agent.

The VR installation *Enter The Hindenburg*, which was developed by the authors in the context of the "OpenCulture-Meets-BW"-Hackathon, provides a good basis for the intended experiment. The intention of *Enter the Hindenburg* is to enable participants to experience and explore the historic airship and intangible cultural heritage "Hindenburg". Using the application, participants can explore the airship from different perspectives inside the hangar and get articulated information about the historic context. Besides this passive option, the application allows the user to *interactively* participate. The user is guided by the agent through spoken commands to control a crane and help workers build the airship, and is also given instructions on how to repair and paint the outer skin of the plane.



(a) Disembodied (b) Object

(c) Anthropomorphic Object (d) Human

Fig. 1. Different visual agent representations. (Color figure online)

4.1 Agents

To provide an appropriate setup, we have limited our investigations to four types of agent-mediated interfaces (see Fig. 1), namely:

- disembodied (audio-only without embodiment),
- as an object,
- as an anthropomorphic object,
- and with a human appearance.

The agents only expressed themselves verbally. For better comparability, object manipulation was not demonstrated even though the human avatar would have been able to. The three different variations of agent visualizations used in our evaluation fulfill different, specific needs and support different expectations as discussed in Sect. 1. In addition to conceptual considerations such as the need for high realism in certain virtual training scenarios, technical limitations (e.g., lack of display capabilities in voice assistants such as Amazon's Alexa) and cost (e.g., high vs. low polygonal models) may also influence design decisions (Table 1).

All investigated agents share the same audio track and behavior and only differ in visual appearance and facial animations.

Table 1. Visual features of the different agents.

	Disembodied	Object	Anthropomorphic object	Human
Idle animation	✗	✓	✓	✓
3D-Audio	✗	✓	✓	✓
Gaze tracking	✗	✗	✓	✓
Lip-sync	✗	✗	✓	✓
Gestures	✗	✗	✗	✓

Disembodied (audio-only). Unlike all other agent types studied in this context, the *disembodied* agent has no visual representation and only the environment is visible (see Fig. 1a). the agent's voice is designed in such a way that no spatial orientation can be derived from it.

Object. In virtual worlds, any object can take the place of a protagonist. According to the *media equation*, it can be said that under certain circumstances objects are perceived as communicative objects [20]. Probably the most famous object protagonist is the Pixar *Luxo Jr.* mascot—it demonstrates well how expressive objects can be.

The object has to fit well with the narrative and the environment, so we decided for a shrunken, cartoon version of a Zeppelin (see Fig. 1b). To bring the protagonist to life, we added basic animations: The object moves slightly during the scene and always rotates to face the player, but with a 20° offset so that the user can also see the side of the Zeppelin. Whenever the Zeppelin is in motion (by turning or changing position), the rotors of the engines also turn, relative to speed. The voice was spatially located at the protagonist's location.

Anthropomorphic Object. By adding facial features and other body parts such as arms or legs, any object can be humanized. These types of objects, called anthropomorphic objects, potentially enhance the representation of emotions and information through gesturing or gaze direction [17]. A good example is given in the movie *Cars* by *Pixar*.

Since eyes, and the resulting eye contact, are particularly important for communication, it is therefore also of interest to evaluate a humanized version of our protagonist [2]. We got inspired by the radio character in the movie *The Brave Little Toaster* and implement a humanized version as can be seen in Fig. 1c.

In addition to the animations used in the previously described case, the agent of the anthropomorphic object blinks and the lips move in sync with what is being said.

Human. The final representation is human. To fit the narration, we chose a male person, about 40 years old, wearing a blue shirt, beige corduroy pants, and a gray beret (see Fig. 1d). The body and face were fully rigged and visemes

Fig. 2. Study participant during the "repairing the outer skin" task

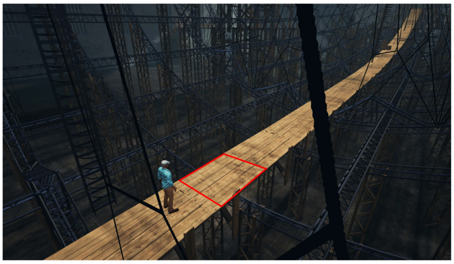
were added to allow lip sync. Procedural animations were also added so that the agent always turns his head to have eye contact with the player. If the user would move so far that turning the character's head is not enough, the character also rotates around its axis in 90-degree increments. Additionally, the character blinks at random intervals. While speaking, the character moves with generic gestures overlaid with default animations. This allows the character to rotate, blink or point its head in the direction of the player while gesturing.

4.2 Tasks of the User

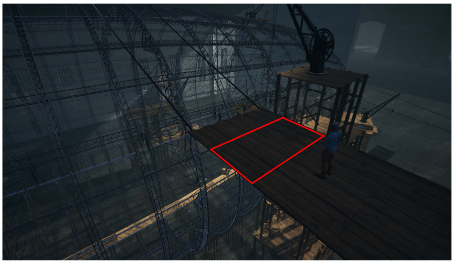
To be able to evaluate the avatars in different use cases, we implemented four tasks, one of which is passive and three are active. As a VR Headset, we used an HTC Vive Pro Eye in combination with the HTC Vive controllers (see Fig. 2).

In the *passive scene*, the user does not have a specific task, but is given information about the *Hindenburg* and its construction at three different positions (see Fig. 3). Once an agent has finished his explanation, he remains in "idle state" until the player presses the trigger button of the VR controller, which teleports the player to the next location.

In the *first interactive scene* (see Fig. 4), the user is instructed to use the levers to control a crane located on the ceiling. The levers are operated with the virtual hands of the player, which he activates by pressing the trigger button. The crane is to be used to move the aluminum struts, which are highlighted in blue, to the target position marked in yellow (see Fig. 4a). Each lever controls one axis (forward/backward, left/right or up/down) of the crane, multiple levers can be operated simultaneously. If the hook of the crane is close to the pallet, it automatically hooks and follows the position of the crane from now on. As soon as the pallet is near the target position, it automatically snaps to the correct place and the task is solved. For increased visibility even in darker areas (see Fig. 4b), two lamps, one shining towards the ceiling and the other being an



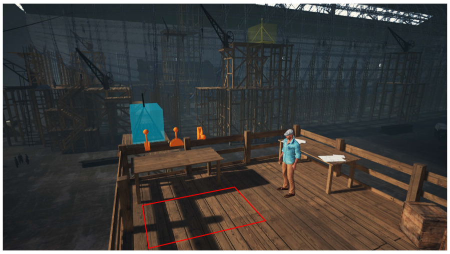
(a) First position



(b) Second position

(c) Third position

Fig. 3. Positions during the *passive scene*. The walkable area is shown in red. Please note that all images show the human agent but have been evaluated for all other agent types. (Color figure online)



(a) Position of the user. The walkable area is shown in red.

(b) Crane, which is controlled by the player

Fig. 4. *Interactive scene one*: controlling a crane in which the player has to bring the aluminum struts (highlighted in blue) to the marked target position (highlighted in yellow). (Color figure online)

omnidirectional beacon with periodic changes in brightness, are mounted on the crane's hook.

Interactive scene two requires the user to repair a part of the outer skin of the airship (see Fig. 5). The outer skin is placed on a table right in front of the user (see Fig. 5a). The tools required for the repair are available on a second table positioned to the right of the user (see Fig. 5b). A yellow light indicates

(a) Position of the user. The walkable area is shown in red.

(b) Repair utensils

Fig. 5. *Interactive scene two*: repairing the outer skin (Color figure online)

(a) Position of the user. The walkable area is shown in red.

(b) Utensils for painting

Fig. 6. *Interactive scene three*: painting the outer skin

where the repair utensils needs to be placed. Once the player is near the target position with a correct component, the placeholder location glows blue and the component can be released to snap into place. At the beginning of the scene, only the fabric is lit. After the fabric has been placed on the outer skin, the player is instructed to attach the fixing clips.

As a *third interactive scene*, the user needs to paint the Hindenburg's outer skin (see Fig. 6). After fixing the hole in the skin, it still remains on the table (see Fig. 6a) so that in can be painted accordingly. Once again, the tools needed are placed on the table right next to the participant (see Fig. 6b). In order to paint the outer skin, all the participant has to do is pass the brush over the outer skin. Once 15% of the outer skin has been painted, the task is considered complete.

5 Study Design

To obtain comparable results in the areas of agent *presence*, *attractiveness*, and *sense of agency*, we relied on established questionnaires. Each item of the used questionnaires was rated on a Likert scale between 1 and 6 and later combined as described in the subsequent subsections.

5.1 Presence

To measure the perceived *presence*, the *iGroup presence questionnaire*[1] [22] was used. This questionnaire measures *presence* based on the four components *spatial presence*, *involvement*, *experienced realism* (*realism*), and *general presence*. While *spatial presence* refers to the spatial location in the virtual world, *involvement* refers to the degree of influence the user has in the virtual world. *Realism*, on the other hand, describes to which degree the virtual world resembles the real world. In addition to the three components mentioned above, there is one item that loads on all three factors ("In the computer-generated world, I had the impression of having been there") and is therefore rated as *general presence*. Since different items load differently on a component, the value ranges of the individual components are different for the respective items. The aggregated component *spatial presence* has a value range between 4.1 and 24.5, *involvement* between 3.2 and 19.0, *realism* between 3.4 and 20.1, and *general presence* between 1 and 6. These aggregats are calculated by the sum of all products of the chosen value by the participant and the corresponding factor loading.

5.2 Attractiveness

To measure *attractiveness* the *AttrakDiff 2* questionnaire by Hassenzahl et al. [8] is used. The measurement of a product's *attractiveness* is thereby separated in three categories: *hedonic quality stimulation* (*stimulation*), *hedonic quality identity* (*identity*), and *pragmatic quality*. If a product is usable for environmental manipulation, Hassenzahl et al. label them as *pragmatic*. When an object extends the possibilities of a user through new functions, this is referred to as *hedonic quality*. While *stimulation* refers to personal development by improving one's own performance, *identity* reflects one's own personality in the product. After all, Hassenzahl et al., developed 21 items. The components were labeled according to the categories of items that loaded particularly high on them. The component *identity* has values in the interval of 5.7 and 34.1, *stimulation* has a range of values between 5.9 and 35.5, and *pragmatic quality* between 5.2 and 31.3.

5.3 Agency

To measure the *sense of agency* the *The Sense of Agency Scale* questionnaire by Tapal et al. [23] is used. It contains thirteen items. Six items relate to *sense of positive agency* and seven to *sense of negative agency*. *Sense of positive agency* describes the perceived degree to which participants felt they initiated the actions, whereas *sense of negative agency* describes the absence of this feeling. Since the items are weighted here as well, the components also have a different lower and upper bound, which is between 3.0 and 18.5 in the case of *sense of positive agency* and between 2.6 and 15.8 for *sense of negative agency*.

[1] http://www.igroup.org/pq/ipq/index.php.

5.4 Procedure

Once the participant arrived at our lab, they were given a short introduction on how to control the application by the study supervisor while calibrating the eye-tracking system. As soon as the participant was ready, the first scene was started (exemplarily shown in Fig. 2)[2]. All scenes had to be completed without the help of the supervisor, only in case of multiple misinterpretation or complete misunderstanding of the task the supervisor gave a hint to solve the problem. When the participant successfully completed the given task, they were teleported back to the neutral initial scene (blue sky, white ground). The participants were asked to take off the glasses, with the supervisor assisting them, and to answer the questionnaire on a laptop. This procedure was then repeated to a maximum of four iterations, depending on the respondent's condition, so that each respondent went through at least two and a maximum of four scenes. When selecting the possible combinations, a 4×4 matrix was used to ensure a precisely balanced distribution. Completing all 4 tasks including the onboarding and eyetracking calibration process took approximately 1 h.

5.5 Participants

34 people participated in the study, 21 female and 13 male. Age ranged from 19 to 52 years (mean $M = 29.41, sd = 8.59$). The VR experience, which was assessed participatively on a scale from 1 (no experience) to 6 (regular experience), has a mean of $m = 2.47$ and a standard deviation of $sd = 1.24$. The fact that many of the participants tested all four scenes, results in a total number of 112 observations, 28 observations per group.

6 Results

In the following, the user study is evaluated to confirm or reject the hypotheses as stated in Sect. 3. For this purpose, all components between the four groups *disembodied*, *object*, *anthropomorphic object* and *human* are analyzed for significance using the *Kruskal-Wallis Test*, since the data is not normally distributed [19]. Table 2 presents a normalized summary of all results.

6.1 Presence

As presented in Table 3 all four measures related to presence, namely *spatial presence*, *involvement*, *realism*, and *general presence* show no significant difference between the four visualizations. Therefore, the hypothesis **H1** [***Presence*** *is influenced by the type of visual appearance of the agent*] <u>cannot</u> be confirmed.

[2] A screen-capture of an exemplary study situation is provided on youtube. https://youtu.be/dD7RQ2inWdk.

Table 2. Since the value ranges of all comp/onents vary due to different factor loadings, the values were scaled to fit between 1 and 6. Unscaled values can be found in the corresponding sections. The given values correspond to the mean value of the results. Superscript numbers ^{1,2,3,4} mark significant differences ($p < 0.05$) between the items within a row.

	Disembodied (1)	Object (2)	anthro. Object (3)	Human (4)
Presence				
Spatial presence	5.54	5.06	5.45	5.19
Involvement	4.79	4.52	4.64	4.39
Realism	4.66	4.27	4.61	4.23
General presence	5.18	4.89	5.21	4.93
Attractiveness				
Identity	3.89^{4}	3.76^{4}	4.19	$4.39^{1,2}$
Stimulation	$3.41^{2,3}$	4.06^{1}	4.27^{1}	3.86
Pragmatic quality	4.06^{4}	$3.81^{3,4}$	4.43^{2}	$4.50^{1,2}$
Sense of agency				
Negative *agency*	2.44	2.40	2.20	2.27
Positive *agency*	5.16	4.87	5.31	4.92

Table 3. Results for *presence* components. The values shown in the table correspond to the mean value and standard deviation of the results.

	Disembodied (1)	Object (2)	anthro. Object (3)	Human (4)
Spatial presence	22.6 ± 3.57	20.6 ± 4.81	22.3 ± 3.9	21.2 ± 3.21
Involvement	15.2 ± 4.38	14.3 ± 4.82	14.7 ± 4.2	13.9 ± 3.66
Realism	15.6 ± 3.68	14.3 ± 3.79	15.5 ± 3.16	14.2 ± 3.18
General presence	5.18 ± 0.86	4.89 ± 0.87	5.21 ± 0.92	4.93 ± 0.77

6.2 Attractiveness

In Table 4 the influence on attractiveness is presented. It can be seen that there exist several differences between the agent types for all three sub-categories especially in regard to the human agent which shows significant differences with *disembodied* as well as with *object* for *identity* and *pragmatic quality*. In case of *stimulation*, the *disembodied* agent shows the largest differences to the other agents. Here we find significant differences between *disembodied* and *object* as well as *anthropomorphic object* but not for *human*. Therefore, the hypothesis **H2** [**Attractiveness** is influenced by the type of visual appearance of the agent] can be confirmed.

Table 4. Results for the *attractiveness* components. The values shown in the table correspond to the mean value and standard deviation of the results. Superscript numbers ^{1,2,3,4} mark significant differences ($p < 0.05$) between the items within a row.

	Disembodied (1)	Object (2)	anthro. object (3)	Human (4)
Identity	22.1 ± 5.03^{4}	21.4 ± 4.09^{4}	23.8 ± 3.38	$25.0 \pm 4.09^{1,2}$
Stimulation	$20.1 \pm 4.61^{2,3}$	24.0 ± 3.07^{1}	25.2 ± 4.42^{1}	22.8 ± 3.07
Pragmatic quality	21.1 ± 3.71^{4}	$19.9 \pm 3.73^{3,4}$	23.1 ± 2.49^{2}	$23.4 \pm 3.03^{1,2}$

6.3 Sense of Agency

By comparing the numbers presented in Table 5 no significant differences are present between all four agent styles: for the component *sense of positive agency*, minimally higher means show up in the groups *human* and *anthropomorphic object* compared to the groups *disembodied* and *object*. For the component *sense of negative agency* the differences are even smaller, only the component *object* has a minimally higher value than the remaining groups. It should also be noted that this component is inversely coded, as it reflects negativity. Thus, a high numerical value stands for the absence of the *sense of agency*. Therefore, the hypothesis **H3** [***Sense of agency*** *is influenced by the type of visual appearance of the agent*] cannot be confirmed.

Table 5. Results for the *sense of agency* components. The values shown in the table correspond to the mean value and standard deviation of the results

	Disembodied (1)	Object (2)	anthro. Object (3)	Human (4)
Negative *agency*	3.48 ± 4.44	3.26 ± 3.48	2.22 ± 3.33	2.57 ± 3.56
Positive *agency*	15.9 ± 3.56	15.0 ± 3.96	16.4 ± 2.98	15.1 ± 4.12

6.4 Further Notes

It should be noted that as a supervisor one could observe in some cases that the non-humanoid Zeppelin was not perceived as the origin of the voice and accordingly not as a communication partner. Other findings that came to the attention of the test supervisor during the execution of the experiments are, for example, that the non-humanized Zeppelin was often not perceived as a communication partner. According to some participants, they did not know that the voice was coming from the Zeppelin. Furthermore, the participants noted that the non-humanized Zeppelin would not fit into the overall scenery due to its comic style, whereas the Zeppelin with humanizations, on the other hand, was rather perceived as fitting, although—except for eyes, ears and mouth—the model is identical. Furthermore, it was noticed that the participants always searched for a visible figure with the *disembodied* advisor, even if the *disembodied* variant was tested first.

6.5 Scene Comparison

The question whether a specific agent type is preferable for either an *interactive scene* or a *passive scene*, was analyzed by taking a specific look at the results for each group. All scenes were compared pairwise for all agents with a *Kruskal-Wallis Test* and did not yield significant differences between any constellation. This means that no implemented agent is preferable for *interactive-* or *passive scenes*.

7 Discussion and Outlook

In this work, we evaluated the influence of agents' appearance in VR on *presence*, *attractiveness*, and the *sense of agency* using existing, established questionnaires. We implemented four different agent styles that all share the same audio-track but differ in appearance. With the presented results we can state that, while differences exist in measured *attractiveness*, no differences in *presence* and *sense of agency* between the different visualizations were found (see Table 2). Further breaking down of the results show that for *all variables measured* there is no significant difference between the *human* and the *anthropomorphic object* agent and only one significant difference between *anthropomorphic object* and *object* can be confirmed. In regard to the *anthropomorphic object* and the *object* comparison, Koda [12], found that—even though the agents' perceived *intelligence* was not influenced by whether a face was added or not— the *engagement* with the agent that had a face was significant larger. So, depending on the context, especially in entertainment and also for learning environments, adding human properties to the agent is a valuable contribution.

From this, we conclude that with current representation possibilities of virtual humans, the agent can be exchanged with an *anthropomorphic object* without losses even in cases of *attractiveness*. The other agents also do not differ in case of *presence* and *sense of agency*. This allows for several possibilities such as framing a teaching or guidance situation with the appearance of an *anthropomorphic agent* that sets up a mental model and conveys the feeling of competency (anthropomorphic saxophone guiding through music exhibition, anthropomorphic calculator giving math classes.).

There are also several arguments why a *disembodied* agent might be a valid choice since it shows no significant difference in case of *presence* as well as in *sense of agency*. For example the fact as stated by McDonnell et al. [14] that audio-only environments offer better concentration on what is spoken might be of advantage when only spoken content is of interest. It is worth mentioning that—especially with the *object* with no humanizations—no spatial localization by audio was possible, sometimes the workers (NPCs) were wrongly identified as interlocutors.

The increasing use of intelligent voice assistants (such as Amazon's Alexa) could lead to an extended acceptance of assistants in non-human appearance. Just recently, Amazon released a new version of their Alexa Echo for kids which comes with a cover to look like a panda bear or similar[3]. With regard to these findings, a non-humanoid proxy (in this case a Zeppelin) is not necessarily a worse choice than a *human*, as long as it is made clear to the user from whom he/she is receiving the information or instructions.

In case of objects, this can be accomplished by adding human features, e.g. a mouth that moves to match the speech. Such extensions can even work in the real world, as evidenced by a study by Ohmura et al. comparing an anthropomorphic version of a printer [18].

The study shows that further research in this area might be of value since for this work, only one specific cartoonish stylized *object* was humanized. We see that, with upcoming technological advances, the representation of the virtual *human* will be even closer to what a real human looks like.

References

1. Banks, J., Bowman, N.D.: Emotion, anthropomorphism, realism, control: validation of a merged metric for player–avatar interaction (PAX). Comput. Hum. Behav. **54**, 215–223 (2016). https://doi.org/10.1016/j.chb.2015.07.030, https://www.sciencedirect.com/science/article/pii/S0747563215300406
2. Bayliss, A., Tipper, S.: Predictive gaze cues and personality judgments: should eye trust you? Psychol. Sci. **17**, 514–20 (2006). https://doi.org/10.1111/j.1467-9280.2006.01737.x
3. Belanche, D., Casaló Ariño, L., Flavian, C.: Artificial intelligence in fintech: understanding robo-advisors adoption among customers. Ind. Manag. Data Syst. **119**, 1411–1430 (2019). https://doi.org/10.1108/IMDS-08-2018-0368
4. Bergmann, K., Eyssel, F., Kopp, S.: A second chance to make a first impression? How appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time. In: Nakano, Y., Neff, M., Paiva, A., Walker, M. (eds.) IVA 2012. LNCS (LNAI), vol. 7502, pp. 126–138. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33197-8_13
5. Cherif, E., Lemoine, J.-F.: Human vs. synthetic recommendation agents' voice: the effects on consumer reactions. In: Rossi, P. (ed.) Marketing at the Confluence between Entertainment and Analytics. DMSPAMS, pp. 301–310. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-47331-4_53
6. Forlizzi, J., Zimmerman, J., Mancuso, V., Kwak, S.: How interface agents affect interaction between humans and computers. In: Proceedings of the 2007 Conference on Designing Pleasurable Products and Interfaces, DPPI 2007, pp. 209–221. Association for Computing Machinery, New York (2007). https://doi.org/10.1145/1314161.1314180
7. Gulz, A., Haake, M.: Social and visual style in virtual pedagogical agents. In: Workshop on Adapting the Interaction Style to Affective Factors associated with the 10th International Conference on User Modeling (2005)

[3] https://www.theverge.com/2021/6/29/22554428/amazon-reading-sidekick-alexa-echo-skill-kids-voice-profiles.

8. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität (2003)
9. Hepperle, D., Ödell, H., Wölfel, M.: Differences in the uncanny valley between head-mounted displays and monitors. In: 2020 International Conference on Cyberworlds (CW). IEEE (2020). https://doi.org/10.1109/cw49994.2020.00014
10. Hepperle, D., Purps, C.F., Deuchler, J., Wölfel, M.: Aspects of visual avatar appearance: self-representation, display type, and uncanny valley. Vis. Comput. 1–18 (2021). https://doi.org/10.1007/s00371-021-02151-0
11. Hepperle, D., Weiß, Y., Siess, A., Wölfel, M.: 2D, 3D or speech? A case study on which user interface is preferable for what kind of object interaction in immersive virtual reality. **82**, 321–331 (2019). https://doi.org/10.1016/j.cag.2019.06.003
12. Koda, T., Maes, P.: Agents with faces: the effect of personification. In: Proceedings 5th IEEE International Workshop on Robot and Human Communication, RO-MAN 1996 TSUKUBA, pp. 189–194. IEEE (1996). https://doi.org/10.1109/ROMAN.1996.568812
13. Lee, H., et al.: Annotation vs. Virtual tutor: comparative analysis on the effectiveness of visual instructions in immersive virtual reality. In: 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) (2019). https://doi.org/10.1109/ISMAR.2019.00030
14. McDonnell, R., Breidt, M., Bülthoff, H.H.: Render me real? Investigating the effect of render style on the perception of animated virtual humans. ACM Trans. Graph. **31**(4) (2012). https://doi.org/10.1145/2185520.2185587
15. Miao, F., Kozlenkova, I.V., Wang, H., Xie, T., Palmatier, R.W.: An emerging theory of avatar marketing. J. Mark. (2021). https://doi.org/10.1177/0022242921996646
16. Niewiadomski, R., Demeure, V., Pelachaud, C.: Warmth, competence, believability and virtual agents. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) IVA 2010. LNCS (LNAI), vol. 6356, pp. 272–285. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15892-6_29
17. Osawa, H., Ohmura, R., Imai, M.: Embodiment of an agent by anthropomorphization of a common object, vol. 10, pp. 484–490 (2008). https://doi.org/10.1109/WIIAT.2008.129
18. Osawa, H., Ohmura, R., Imai, M.: Embodiment of an agent by anthropomorphization of a common object. In: 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 2, pp. 484–490 (2008). https://doi.org/10.1109/WIIAT.2008.129
19. Ostertagova, E., Ostertag, O., Kováč, J.: Methodology and application of the Kruskal-Wallis test. Appl. Mech. Mater. **611**, 115–120 (2014)
20. Reeves, B., Nass, C.: The media equation: how people treat computers, television, and new media like real people and places. Bibliovault OAI Repository, the University of Chicago Press (1996)
21. Roth, D., et al.: Avatar realism and social interaction quality in virtual reality. In: 2016 IEEE Virtual Reality (VR), pp. 277–278 (2016). https://doi.org/10.1109/VR.2016.7504761
22. Schubert, T., Friedmann, F., Regenbrecht, H.: The experience of presence: factor analytic insights. Presence **10**, 266–281 (2001). https://doi.org/10.1162/105474601300343603
23. Tapal, A., Oren, E., Dar, R., Eitam, B.: The sense of agency scale: a measure of consciously perceived control over one's mind, body, and the immediate environment. Front. Psychol. **8**, 1552 (2017). https://doi.org/10.3389/fpsyg.2017.01552

24. Torre, I., Carrigan, E., McDonnell, R., Domijan, K., McCabe, K., Harte, N.: The effect of multimodal emotional expression and agent appearance on trust in human-agent interaction. In: Motion, Interaction and Games, MIG 2019. Association for Computing Machinery, New York (2019). https://doi.org/10.1145/3359566.3360065
25. Wölfel, M.: Acceptance of dynamic feedback to poor sitting habits by anthropomorphic objects. ACM (2017). https://doi.org/10.1145/3154862.3154928