Utrecht University (2022)

Thesis paper submitted in partial fulfillment of

Master of Science in
Game and Media Technology

Investigating the Effects of Appearance and Animation Realism on the Perception of Expressive Photorealistic Virtual Humans

Submitted by: Nabila Amadou,
First Supervisor: Dr. Zerrin Yumak,
Second Supervisor: Prof. Dr. Albert Ali Salah,
Daily Supervisor: PhD Canditate Kazi Injamamul Haque

Utrecht University, Netherlands

ARTICLE INFO

Article history:
Received 25 September 2022
Received in final form 09 October 2022

Keywords:
Virtual humans
Perception
Animation realism
Appearance realism

ABSTRACT

Currently, digital assistants are predominantly limited to the modality of audio or text. As technology is moving at a rapid pace however, we can soon expect them to adopt full artificial human appearances. Investigating the perception of these virtual humans can help guide designers to better engineer these artificial digital humans that are to take on the roles of digital companions.
In the past, numerous research studies have investigated the perception of virtual humans under the lens of appearance realism and animation realism among other characteristics. However, virtual humans formerly termed photorealistic, cannot be called as such under today's standards anymore. While previously, it was a well-established fact that photorealistic characters elicit the uncanny valley phenomenon - recent research investigations have challenged this by re-opening research questions and making use of new state-of-the-art photorealistic digital human models. This paper aims to contribute towards this research effort by investigating main and interaction effects of appearance realism and animation realism for expressive virtual humans.
In a study with 62 participants, there were statistically significant effects of appearance realism for social presence. Additionally, animation realism as well as appearance realism played a significant effect on perceived discomfort towards the virtual human. Finally, evidence hints at higher animation realism being received significantly better when paired with photorealistic characters.

1. Introduction

As we are heading into a more digitized world, the way humans interface with computers is set to transform itself. It's already possible to talk to computer assistants through sophisticated voice interfaces, such as Amazon's Alexa [1] or Google Assistant [2]. Among the next steps is to add to the virtual assistant's audio with visuals. In media, whether it's games, film or in future every day applications with virtual human assistants, we will be faced with these human-like computerized entities that look more and more realistic. Therefore, it is important to investigate how we perceive virtual humans and further uncover the reasons why we perceive virtual humans in certain ways. That way industry developers and artists can better plan for appropriate virtual human designs with regards to appearance, be-

havior and animation quality. It follows that, in recent years, the research field investigating the perception of virtual humans has gained a lot more traction. Especially, as virtual humans have become incredibly realistic, to a point where they closely resembled real humans. Specifically, the uncanny valley has accrued a significant amount of attention - a phenomenon where artificial humans are seen as unappealing when their realism is very close to the real counterpart. Small imperfections would trigger aversion behavior in observers. A lot of research has been done to investigate which elements about virtual agents create unappealing effects - the research in question however tended to generally focus on virtual humans that aren't photorealistic in today's standards. Recent industry breakthroughs - for instance with Epic Games' MetaHumans [3], Ziva Dynamic's [4] or Reblika's virtual humans [5] have achieved a level of quality in which virtual humans are indeed almost indistinguishable from real humans. New insights might overturn conclusions formerly drawn from old studies with virtual character appearances that can be termed semi-realistic by today's standards.

Previous research has suggested that higher gesture motion realism [6], higher appearance realism [7] and higher facial animation realism in virtual humans [8] lead to lower discomfort. Comparisons have been made between photorealistic and stylized human agents [6], real humans and photorealistic agents [9] as well as photorealistic with semi-realistic virtual humans [7].

Little research has been published to investigate interactive effects of manipulation of appearance realism and facial animation realism of human agents in a capacity that exceeds eye-gaze and eye-blinking manipulation [10]. Exploring these channels with an emotionally expressive virtual agent is important in identifying in how far compromising on facial animation realism and appearance realism can impact the discomfort felt towards the virtual human.

In this paper, main and interaction effects of appearance realism and facial animation realism on the perception of a virtual human and within the context of emotional scenarios are investigated. I searched for evidence with regards to how the simultaneous manipulation of appearance realism and facial animation realism could impact the social presence of and discomfort towards an emotionally expressive virtual human. In an on-screen user study with 62 participants, their effects on perceived emotion intensity, social presence as well as perceived appeal and eeriness among other characteristics were analyzed. Social presence was included as a measure in this study, alongside affinity and perceived realism measures.

2. Related work

At the center of this research lies the investigation of social presence and affinity of a virtual human while appearance realism, animation realism and emotion context are manipulated. As social presence and affinity are by themselves rather abstract terms, it is important to explain the terms and their use in this paper.

Thus, this section is organized as follows: In a first part, the independent variables of this study are highlighted. In a second part, the dependent variables are explored as well to give a complete overview of the variables as used in this study.

2.1. Part I - Manipulating appearance realism, animation realism and emotion context

The question arises whether animation realism mismatches with the appearance realism of a virtual human can induce discomfort.

There is little research involving the simultaneous manipulation of a virtual human's appearance and animation realism. However, there are still a number of notable publications to be mentioned in this context.

Ferstl et al. [6] investigated appearance and gesture motion realism of social agents. Here they found that the motion condition with highest human-like realism was significantly more likable than all other reduced motion conditions. For appearance realism investigations, they only compared between a human and an anthropomorphic robot model. Interestingly, the robot model was perceived as significantly more likable than the virtual human, also for the most realistic, human-like gesture motion condition.

An older study, McDonnell et al. [10] investigated the effect of different render styles and animation anomalies on the perception of virtual humans. Two of the abstract cartoon renders as well as the most realistic renders were considered significantly more appealing. It was the render styles categorized between realistic and cartoonish that were perceived as the least appealing. When testing motion anomalies by removing eye-blinking and eye-gaze or turning motion off for half of the face, they found the anomalies were perceived a lot more unpleasant in the case of an ill looking realistic human. On the other hand, a cartoonish render was perceived as the most appealing. These results hint that more realistic virtual humans make people more sensitive towards imperfections. It is to be noted however, that neither of these studies so far investigated emotionally expressive virtual humans - a gap this study aims to contribute towards.

In addition, there are a number of works either investigating the isolated manipulation of appearance realism or animation realism and their effects on the perception of virtual agents.

Notably, regarding investigating animation realism, Tinwell et al. [8] researched if there could be a correlation between traits associated with psychopathy and perceptions of the uncanny. They removed any eyebrow motion in virtual characters - effectively removing an upper face startle response - and found significant correlation between eeriness ratings with full animation realism and reduced upper face motion.

Finally, a number of studies have investigated the perception of emotionally expressive photorealistic virtual humans. Of relevance are recent papers that used virtual humans that are more realistic, such as in this study. These studies include a paper by Zibrek et al. [11] as well as a research endeavor by Higgins et al. [7]. In both studies, appearance realism was manipulated for an expressive virtual human. In the case of Higgins et al., Epic Games' MetaHumans were used, with same appearance realism conditions as in this paper. Both of these studies had conclusions which suggested that the uncanny valley with today's

photorealism standard has been crossed, as more photorealistic renders were perceived as more appealing in user studies.
This research was inspired with these results in mind, as it is interesting to look for more evidence to support or challenge their results. In addition, observing conjunction effects of appearance realism and facial animation realism, in a similar fashion to McDonnell et al.'s work could lead to renewed insights with virtual humans reaching a new state-of-the-art photorealism standard.

2.2. Part II - Measuring affinity, social presence and emotion intensity

The Uncanny Valley

Research involving the perception of realistic virtual humans is always accompanied by discussions about the uncanny valley. Originally, its concept was introduced by Mori [12] in the context of robotics. Researchers discovered that it applied to virtual humans as well [13]. The uncanny valley describes that with higher appearance realism, the appeal of virtual humans increases. However, at a point where they are very close to looking like realistic humans, they trigger instincts of aversion. Many research investigations made it their goal to find out possible reasons for the perception of the uncanny valley. Some studies suggested that a mismatch of facial features triggers the uncanny valley effect, such as eyes that were too big [14]. Others suggested that a cross-modal realism mismatch would lead to the uncanny valley effect [15].
Towards the goal of quantifying the effect of the uncanny valley, researchers introduced measures for affinity towards the virtual human. Among those are likability or appeal [10] [9], eeriness [16], familiarity [10] and attractiveness [17] of the virtual human.

Social presence

While the term 'social presence' was coined and is oftentimes used in VR application contexts, it is nevertheless in its original definition described as when one perceives "that a form, behavior, or sensory experience indicates the presence of another intelligence" as stated by Frank Biocca [18]. This term, social presence, later robustly defined to describing the perception of a "being with another" by Biocca et al. [19] is transferable to other interaction contexts, such as the on-screen experiment in this paper. Many studies have investigated the perception of virtual humans under the lens of social presence.
For instance, a recent study by Guimaraes et al. [20] has found evidence for social presence felt towards virtual agents being higher in VR when compared to a screen-only interaction. Note that this conclusion is to be regarded within the context of their experiment as they didn't compare multiple virtual humans with different levels of appearance realism. Other studies, by Zibrek and McDonnell [16] and Zibrek et al. [11], performed investigations into how different rendering styles influence social presence in an immersive environment and found a significant effect of appearance realism as well. While evidence suggests that higher social presence can be achieved in VR and while this hints at a potential interaction effect between VR and appearance realism, this is outside the scope of this paper. This study aims at investigating the effect of photorealistic humans on social presence by using Epic Games' MetaHumans. As the highest level of detail possible for MetaHumans didn't run reliably enough within a real-time VR environment, this study was limited to an on-screen experiment.
A different research endeavor, by Higgins et al. [15], found evidence suggesting that multimodal congruence increases social presence with regards to audio and visual appearance of the virtual human. While these research insights hint at a potential interaction effect between a virtual human's appearance realism and their voice on social presence, it is outside of the scope of this paper.
Essentially, it becomes clear that social presence has established itself as an important measure for evaluating how believable a virtual agent is as a sentient human.

Emotion Intensity

Similar research investigations to this study have looked into the emotional response induced when faced with virtual humans, such as [16] and [20]. In this research endeavor however, I decided to take a step back and investigate whether the displayed emotion stimuli were perceived as intended in the first place. Similar to [21], the perceived emotion intensity for all displayed emotions were measured. This would help identify if the emotion scenarios would be perceived as designed.

2.3. Contribution of the paper

The contribution of this paper lies in the investigation of main and compounding effects of animation realism as well as appearance realism in an emotionally expressive virtual human. While previous papers investigated the main effects of appearance realism and animation realism as previously discussed, none have investigated interaction effects within the context of emotionally expressive virtual humans. Additionally, it is interesting to study how photorealistic characters of today - here represented by Epic Games' MetaHumans - can challenge past research insights regarding the uncanny valley.

3. Experiment Design

To investigate the perception towards a virtual human and how it changes based on appearance realism and animation realism within different emotional scenario contexts, three independent variables were formally introduced as shown in Table 1. Three investigations were opened and I formulated the following hypotheses:

3.1. Investigation 1: Main effects of animation realism

- **H1: Full animation realism will lead to significantly higher social presence with the virtual human.** I expect upper face motion to be important for more natural and human-like virtual agents. Therefore, the sense of being in the presence of a sentient human being will increase.

- I expect virtual humans with lacking upper face motion to induce feelings of discomfort. Therefore, higher facial

Table 1. Independent variables of this study

Independent variable	Levels	Level description
Appearance Realism	Photorealistic	Current photorealism standard
	Semi-realistic	Former photorealism standard
Animation Realism	Full	Full face animation
	Medium	No eyebrow animation
	Low	No upper face animation
Emotion Scenario	Neutral	
	Happy	
	Angry	
	Sad	

animation realism will lead to more positive affinity towards the virtual human as the motions will more closely resemble real human movement. Therefore hypotheses are formulated as:
H2a: Full animation realism will lead to significantly higher appeal.

H2b: Full animation realism will lead to significantly higher familiarity.

H2c: Full animation realism will lead to significantly higher attractiveness.

H2d: Low animation realism will lead to significantly higher eeriness perceptions.

- **H3: Full animation realism will lead to more intensely perceived emotions.** I expect that congruence of emotional expressivity in facial animation and voice to be indicative of perceived emotion intensity. With included upper face motion, the facial expressions will be more congruent with the emotional voice-over. Therefore, the emotions in the scenarios will be perceived as more intense.

- I expect lacking upper face motion to lead to significantly decreased realism perceptions. Therefore, hypotheses are formulated as:

  H4a: Full animation realism will lead to higher perceived appearance realism.

  H4b: Full animation realism will lead to higher perceived face movement realism.

  H4c: Full animation realism will lead to higher perceived behavior realism.

  H4d: Full animation realism will lead to higher perceived overall realism.

3.2. Investigation 2: Main effects of appearance realism

- **H1: Photorealistic appearance will lead to significantly higher social presence scores than Semi-realistic appearance.** I expect higher appearance realism to be important for the virtual agent's believability as a sentient human. Therefore, Photorealistic appearance will increase the sense of being in the presence of a sentient human.

- I expect a more human and a more detailed character to be seen as more appealing. A semi-realistic virtual human is known to induce more discomfort as it is representative of former state-of-the-art realism which was considered to be perceived as uncanny. As there has been evidence suggesting that today's photorealism has crossed the uncanny, I expect it to receive significantly higher affinity scores. Hypotheses are thus formulated as:

  H2a: Photorealistic appearance will lead to significantly higher appeal than Semi-realistic appearance.

  H2b: Photorealistic appearance will lead to significantly higher familiarity than Semi-realistic appearance.

  H2c: Photorealistic appearance will lead to significantly higher attractiveness than Semi-realistic appearance.

  H2d: Semi-realistic appearance will lead to significantly more eeriness than Photo-realistic appearance.

- **H3: Photorealistic appearance will lead to more intensely perceived emotions.** I expect that the emotions portrayed by a photorealistic human will be perceived as more real and therefore more intense.

- I expect the higher texture and groom detail of Photorealistic appearance to be perceived as a lot more human-like than the Semi-realistic version.
  H4a: Photorealistic appearance will lead to higher perceived appearance realism than Semi-realistic appearance.

  H4b: Photorealistic appearance will lead to higher perceived face movement realism than Semi-realistic appearance.

  H4c: Photorealistic appearance will lead to higher perceived behavior realism than Semi-realistic appearance.

H4d: Photorealistic appearance will lead to higher perceived overall realism than Semi-realistic appearance.

3.3. Investigation 3: Interaction effects of animation realism and appearance realism

In conjunction with the hypotheses from Investigations 1 and 2, I expect their combined effects to be significant as well.

- **H1: Photorealistic appearance & Full Animation Realism will lead to significantly higher social presence scores.**
- **H2a: Photorealistic appearance & Full Animation Realism will lead to significantly higher appeal.**

  H2b: Photorealistic appearance & Full Animation Realism will lead to significantly higher familiarity.

  H2c: Photorealistic appearance & Full Animation Realism will lead to significantly higher attractiveness.

  H2d: Semi-realistic appearance & Low Animation Realism will lead to significantly more eeriness.

- **H3: Photorealistic appearance & Full Animation Realism will lead to more intensely perceived emotions.**
- **H4a: Photorealistic appearance & Full Animation Realism will lead to higher perceived appearance realism.**

  H4b: Photorealistic appearance & Full Animation Realism will lead to higher perceived face movement realism.

  H4c: Photorealistic appearance & Full Animation Realism will lead to higher perceived behavior realism.

  H4d: Photorealistic appearance & Full Animation Realism will lead to higher perceived overall realism.

3.4. Stimuli creation

The independent variables as highlighted on Table 1 are further described in depth:

Appearance Realism

To investigate the effect of appearance realism, the latest released MetaHuman, Kioko, was used. The character is freely available on the MetaHuman Creator site [3]. This ensured the most realistic MetaHuman with highest quality skin textures and grooms at the time. Two appearance realism conditions would be compared: *Photorealistic* and *Semi-realistic* (see Figure 1). The *Photorealistic* appearance condition was to be represented by the highest quality level of detail possible for MetaHumans, LOD0. The *Semi-realistic* appearance condition was represented by LOD4 of the MetaHuman to ensure comparability with previous research where MetaHuman's LOD0 and LOD4 were compared [7]. The *Semi-Realistic* appearance condition here is representative of what was formerly considered to be the photorealism standard as in Higgins et al [7].

Fig. 1. Virtual human "Kioko" as used in this study, a MetaHuman released by Epic Games. On the left, the *Photorealistic* version (LOD0). On the right, the *Semi-realistic* version (LOD4).

Animation Realism

As the same character was used for both Appearance Realism conditions, it ensured that the available blendshapes would remain the same and that the same control rig could be used to run the animations. This means that animation sequences could remain exactly the same for both versions, allowing for a better base for comparing results.

There were three levels to investigate how facial animation realism could impact the perception of the virtual human (in conjunction with Appearance Realism and in the context of different Emotion Scenarios). As lack of upper face motion has been attributed to higher perceptions of eeriness, it was interesting to investigate this effect for virtual humans as well [8]. Therefore, next to the *Full* Animation Realism condition, the anomaly conditions for removed eyebrow motion (*Medium* Animation Realism) and removed upper face motion (*Low* Animation Realism) were investigated as well.

Emotion Scenarios

Four emotional scenarios were designed: *Neutral, Happy, Angry* and *Sad*. In a previous study, Boaz et al. investigated which sentences most correlated with perceived emotions [22]. This was used as a reference to craft 20-second scenario sequences in which the virtual human was to express these emotions. Per emotion 1-3 sentences from [22] were used.

3.5. Performance-driven animation

Animation sequences for the four emotion scenarios were created with state-of-the-art motion capture technology. I performed for all scenarios with a previously prepared script for the role of a female virtual human. For the facial animation, a Dynamixyz head mounted camera was used to record video sequences. Using Performer2SV [23] and Autodesk Maya [24],

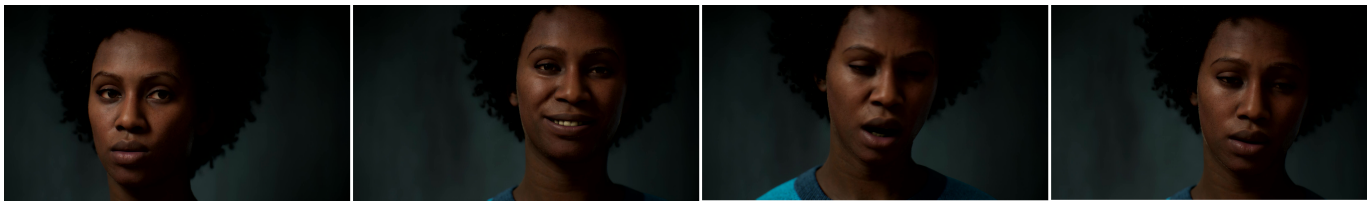


Fig. 2. Screen captures of all Emotion Scenarios (here for the *Photorealistic* appearance condition). From left to right: *Neutral, Happy, Angry, Sad*.

facial expressions were then retargeted onto the MetaHuman.
Although this study focuses on the effects of facial animation realism, it was important to ensure that the body motion of the virtual human was natural - even if the virtual human was visible only from the shoulders up to the experiment participants. Towards that goal, a Vicon motion capture system was used to record natural body movement during the emotional scenario sequences and retarget the motions onto the MetaHuman skeleton.
The scenes were rendered in Unreal Engine 5 with the environment directly taken over from the MetaHuman sample project as provided by Epic Games [3]. This means that lighting as well as background remained unchanged from the Epic Games MetaHuman sample project. Facial animation sequences, body animation sequences as well as the voice recordings were synced up within Unreal Engine to produce the video sequences for the on-screen user study (see Figure 2).

3.6. Measures

For this research, a user study was conducted where participants watched on-screen videos of a virtual human with a view from shoulders up 2. They were guided through the experiment through a survey created on Qualtrics [25]. This survey first displayed a consent form and instructions. Then it would ask demographics questions, informing about gender, age and how much experience they had with virtual humans so far.
After being randomly assigned the *Photorealistic* or the *Semi-realistic* appearance condition, participants were then presented with randomly ordered videos displaying different levels of Animation Realism for all Emotion Scenarios. After each video, participants were asked to first answer an attention check question (asking a simple question about what the virtual human was talking about). Then, they answered a categorical question pertaining to which emotion they perceived from the virtual human. Finally, they were presented with statements representing the dependent variables of the study. They asked to rate on a Likert scale from 1 (”Not at all”) to 7 (”Extremely”) how much they agreed with the statements. These statements were taken from previous research studies investigating the perception of virtual humans as they had been tested and validated before. An overview can be found in Table 2.
As social presence was one of the elements to be investigated, the social presence questionnaire by Bailenson et al. [26] was used. Predominantly utilized in the research field investigating the perception of virtual humans and in accordance with a request from the research community to use the same social presence measures for better comparability of studies [27], its questionnaire items were used in this study as well.
The dependent variables pertaining to the group of ”Affinity” would give insights into how appealing, eerie, familiar and attractive participants found the virtual human. While the variables ”Appeal”, ”Eerie” and ”Familiar” originated from the study by McDonnell et al. [10], the variable ”Attractive” they added in a VR study in which they found that attractiveness did play a role in how eerie/appealing a virtual character is perceived [17].
An additional dependent variable is that of ”Intensity” to investigate how intensely participant perceived the virtual human’s emotions from Wisessing et al.’s study [21].
Finally, a number of dependent variables, assigned to the group ”Realism” were measured to rigorously evaluate whether intended experiment effects and the implementation quality was sufficient (see Table 2). Items in this group were taken over from Zibrek et al. [16] where they were previously validated. However, as this study puts a focus on facial animation realism, the item ”Movement Realism” from Zibrek et al.’s study is separated into ”Face Movement Realism” and ”Body Movement Realism”. Additionally, an item, ”Voice Realism” was added to investigate whether the actress delivered a performance with sufficient quality, given that it wasn’t a professional actress.

3.7. Participants

For the user study, participants were recruited on Amazon Mechanical Turk [28]. Participants were reimbursed 4$. A total of 94 people took part in the study, however 32 had to be excluded due a number of failed attention checks.
This means that there were a total of 62 participants, 31 of which were randomly assigned scenario videos for the *Photorealistic* Appearance Realism condition and 31 of which were randomly assigned scenario videos for the *Semi-realistic* Appearance Realism condition.
Each participant watched 12 videos as for every Appearance Realism condition, there were four Emotion Scenarios rendered in three Animation Realism levels. After each video, there was one attention check question, asking a multiple-choice question about elements the virtual human was talking about.
Participants’ survey submissions were deemed admissible if they passed at least 11 out of 12 attention check questions. This was due to one error being attributed to a genuine mistake amidst 11 correctly answered attention checks (e.g. a misclick).
The mean age of participants was 35.58 (SD = 8.76) with the youngest participant being of age 21 and the eldest being of age 56. On average, their experience with virtual humans or gaming media in general was slightly above average. The mean being between categorical inputs ”A moderate amount” and ”A lot”.

Table 2. Dependent variables of this study. Participants could rate these questionnaire items on a Likert scale from 1 ("Not at all") to 7 ("Extremely").

Group	Dependent variable	Questionnaire item
Emotion Perception	Intensity	"How intense was the emotion you observed?"
Social Presence	SP1	"It feels as if I am in the presence of another person in the room with me."
	SP2	"It feels as if the girl is watching me and is aware of my presence."
	SP3	"The thought that the girl isn't real crossed my mind often."
	SP4	"The girl appears to be alive."
	SP5	"The girl is only a computerized image, not a real person."
Affinity	Appeal	"I found the girl appealing, likable."
	Eerie	"I found the girl eerie, creepy."
	Familiar	"I found the girl familiar, I have seen a similar person before."
	Attractive	"I found the girl attractive."
Realism	Appearance Realism Perception	"I found the girl's appearance realistic."
	Face Movement Realism	"I found the girl's facial movements realistic."
	Body Movement Realism	"I found the girl's body movements realistic."
	Behavior Realism	"I found the girl's behavior realistic."
	Voice Realism	"I found the girl's voice realistic."
	Overall Realism	"I found the girl realistic overall."

Gender was balanced with 16 being female and 15 being male in each Appearance condition.

4. Results

The goal was to investigate main and compounding effects of appearance realism (*Photorealistic, Semi-realistic*), animation realism (*Low, Medium, Full*) on the social presence and affinity towards the virtual human portrayed in the context of different emotional scenarios (*Neutral, Happy, Angry, Sad*). The scale measures were *Social Presence, Intensity, Appeal, Eerie, Familiar*, and *Attractive* to investigate the hypotheses specifically. Additional scale measures were *Overall Realism, Face Movement Realism, Body Movement Realism, Behavior Realism* and *Voice Realism* to check proper intended experiment design stimuli effects.

Each measure was to be evaluated individually. Therefore, a mixed three-way ANOVA design was employed with between-subject factor *Appearance Realism* and within-subject factors *Animation Realism* and *Emotion Scenario*.

A subsequent analysis using a two-way mixed ANOVA followed. It was evaluated separately for every emotion scenario with between-subject factor *Appearance Realism* and within-subject factor *Animation Realism*. In cases where a two-way ANOVA resulted in statistically significant results, it is specifically highlighted in the following sections.

The homogeneity of variance assumption was tested using Levene's test, while the assumption of sphericity was checked with

Table 3. Hypothesis Conclusions

Hypothesis	Predicted Effect	Supported by analysis?
Investigation 1		
H1	SOCIAL PRESENCE: Animation Realism	No
H2a	APPEAL: Animation Realism	No
H2b	FAMILIAR: Animation Realism	No
H2c	ATTRACTIVE: Animation Realism	**Yes**
H2d	EERIE: Animation Realism	No
H3	INTENSITY: Animation Realism	**Yes**
H4a	APPEARANCE REALISM PERCEPTION: Animation Realism	No
H4b	FACE MOVEMENT REALISM: Animation Realism	**Yes**
H4c	BEHAVIOR REALISM: Animation Realism	**Yes**
H4d	OVERALL REALISM: Animation Realism	**Yes**
Investigation 2		
H1	SOCIAL PRESENCE: Appearance Realism	**Yes**
H2a	APPEAL: Appearance Realism	No
H2b	FAMILIAR: Appearance Realism	No
H2c	ATTRACTIVE: Appearance Realism	No
H2d	EERIE: Appearance Realism	No
H3	INTENSITY: Appearance Realism	No
H4a	APPEARANCE REALISM PERCEPTION: Appearance Realism	No
H4b	FACE MOVEMENT REALISM: Appearance Realism	No
H4c	BEHAVIOR REALISM: Appearance Realism	No
H4d	OVERALL REALISM: Appearance Realism	No
Investigation 3		
H1	SOCIAL PRESENCE: Appearance Realism x Animation Realism	**Yes**
H2a	APPEAL: Appearance Realism x Animation Realism	No
H2b	FAMILIAR: Appearance Realism x Animation Realism	No
H2c	ATTRACTIVE: Appearance Realism x Animation Realism	No
H2d	EERIE: Appearance Realism x Animation Realism	**No***
H3	INTENSITY: Appearance Realism x Animation Realism	No
H4a	APPEARANCE REALISM PERCEPTION: Appearance Realism x Animation Realism	**Yes**
H4b	FACE MOVEMENT REALISM: Appearance Realism x Animation Realism	No
H4c	BEHAVIOR REALISM: Appearance Realism x Animation Realism	No
H4d	OVERALL REALISM: Appearance Realism x Animation Realism	No

(* = Effect was found, yet different from hypothesis)

Mauchly's test. If Mauchly's test returned significant, the degrees of freedom were adjusted using Huyhn-Feldt estimates. In several cases the homogeneity of variance assumption was violated. Then a non-parametric equivalent for a mixed three-way ANOVA was used in the form of the *nparLD* software package in R [29]. In case of statistically significant effects, non-parametric post-hoc pairwise comparisons with Tukey-HSD contrasts and 95% confidence intervals then followed using the software package *nparcomp* in R [30].
An overview of results can be found in Table 3, 4 and 5. In the following, they are described in more detail.

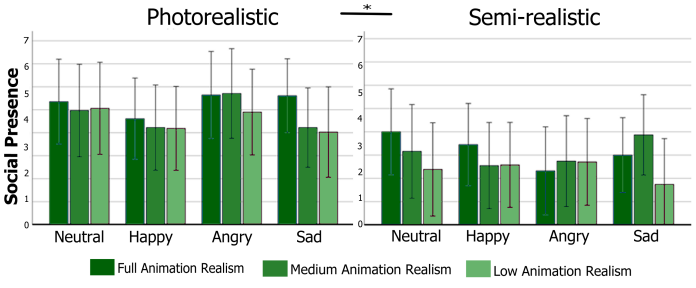


Fig. 3. Estimated marginal means for the variable Social Presence. A significant effect of Appearance Realism was found.

4.1. Social Presence (H1)

Of interest was whether manipulating Animation Realism and Appearance Realism would lead to different perceptions of Social Presence in different emotional scenarios.
As is custom [26], the cumulative score of all social presence items SP1 to SP5 was evaluated after checking the data for reliability (Cronbach's alpha $\alpha = 0.935$). Do note that items SP3 and SP5 contributed inversely to the cumulative scores.
No significant main effect was found for the variable Animation

realism for a 95% confidence interval ($F(2, 120) = 2.716, p = 0.07$). However, the graph of Figure 3 still shows visually distinct effects of animation realism on the Social Presence scores. Figure 3, also displaying the estimated means for *Animation Realism* shows visibly noticeable differences between the three levels, *Low* Animation Realism (estimated $\overline{x} = 3.452$, $SE = 0.559$), *Medium* Animation Realism (estimated $\overline{x} = 3.931$, $SE = 0.547$) and *Full* Animation Realism (estimated $\overline{x} = 4.25$, $SE = 0.499$).

For the independent variable Appearance Realism, a three-way ANOVA revealed a significant main effect ($F(1, 60) = 4.057, p = 0.048$). As seen in Figure 3, there was a noticeably different Social Presence rating for the Appearance condition *Photorealistic* (estimated $\overline{x} = 4.879$, $SE = 0.703$) as compared to the *Semi-realistic* condition (estimated $\overline{x} = 2.876$, $SE = 0.703$), indicating that social presence was indeed increased for the photorealistic appearance condition.

Although no significant interaction effects were revealed in the three-way mixed ANOVA - pairwise comparisons with 95% Bonferroni confidence interval did highlight a significant interaction effect of animation realism with appearance realism. It was overall in the Full Animation Realism condition that Photorealistic Appearance lead to significantly higher social presence scores than Semi-Realistic Appearance ($p = 0.038$).

A separate two-way mixed ANOVA test for each Emotion Scenario was conducted for further investigation. It revealed that only for the *Anger* condition in pairwise comparisons with 95% Bonferroni confidence interval that there was an additional significant interaction effect. Specifically, for the Medium Animation Realism condition, Photorealistic Appearance led to significantly higher Social Presence ratings ($p = 0.035$). Separate tests for *Angry* and *Sad* Emotion Scenarios corroborated the results of the three-way ANOVA - showing significant interaction effects in pairwise comparisons where Full Animation Realism led to significantly higher Social Presence (*Angry*: $p = 0.016$, *Sad*: $p = 0.025$).

Finally, no other interaction effects proved significant for the measure Social Presence.

4.2. Affinity Perceptions (H2)

To investigate was whether Animation Realism or Appearance Realism influenced the affinity measures towards the virtual human in different emotional scenario contexts.

Appeal (H2a)

Neither Animation Realism, nor Appearance Realism had any main effects that proved significant for the measure Appeal. Furthermore, no statistically significant interaction effects were observed.

Overall, the Appeal ratings were above or around average. Photorealistic Appearance (estimated $\overline{x} = 4.223$, $SE = 0.238$) scored slightly higher than Semi-realistic Appearance (estimated $\overline{x} = 4.14$, $SE = 0.238$) on the Appeal scale. Similarly, higher Animation Realism consistently resulted in higher Appeal averages - with Full Animation Realism (estimated $\overline{x} = 4.286$, $SE = 0.179$) being more appealing than Medium Animation Realism (estimated $\overline{x} = 4.19$, $SE = 0.16$) and

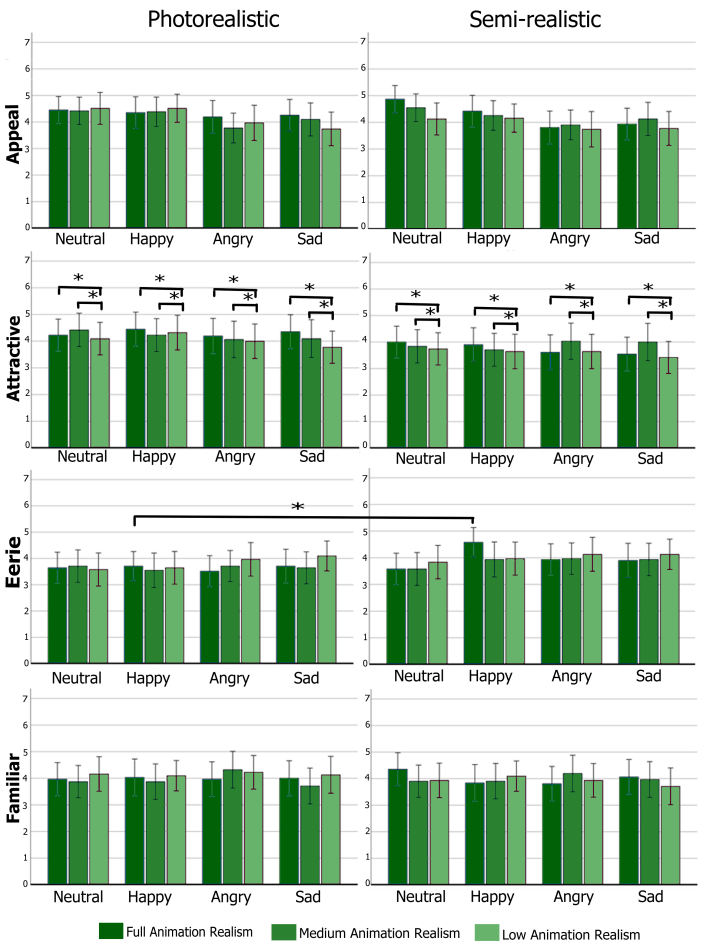


Fig. 4. Estimated marginal means for affinity variables.

Medium Animation Realism being more appealing than Low Animation Realism (estimated $\overline{x} = 4.069$, $SE = 0.19$).

Although, there are no significant effects as related to the hypotheses, a significant effect of Emotion Scenario ($F(2.586, 155.16) = 10.976, p < 0.001$) was observed. Subsequent post hoc pairwise comparisons highlighted that the Emotion Scenarios *Angry* and *Sad* scored significantly lower on the Appeal scale than the Emotion Scenarios *Neutral* and *Happy* (see Table 4) .

Familiar (H2b)

No statistically significant main or interaction effects could be observed for the measure Familiar.

Attractive (H2c)

There was indeed a statistically significant main effect for Animation Realism ($F(1.93, 115.8) = 3.527, p = 0.03$). Subsequent post hoc pairwise comparisons revealed that the *Low* Animation Realism condition scored significantly lower on the Attractive scale than the *Medium* ($p = 0.025$) and *Full* ($p = 0.036$) Animation Realism conditions. Figure 4 highlights this effect, with estimated marginal means for *High* Animation Realism being at an estimated $\overline{x} = 4.036$ ($SE = 0.203$), for

Medium Animation Realism being at an estimated $\overline{x}$ = 4.048 (SE = 0.209) and for *Low* Animation Realism being at an estimated $\overline{x}$ = (SE = 3.831).
No statistically significant main effect of Appearance Realism could be found. However, a general tendency can still be observed in Figure 4. Attractive scores in Photorealistic Appearance reveal an overall higher score (estimated $\overline{x}$ = 4.185, SE = 0.277) than Semi-realistic Appearance (estimated $\overline{x}$ = 3.758, SE = 0.277).
No other statistically significant main or interaction effects could be observed for the measure Attractive.

Eerie (H2d)
A three-way mixed ANOVA yielded the result of no statistically significant main or interaction effects. The outcome safely suggests that the null hypothesis holds which states that the means of all sample sizes are about the same. Due to the complexity of a three-way mixed ANOVA, smaller interaction effects might be obscured however.
In Figure 4, a tendency can be observed for Appearance Realism. Semi-realistic Appearance (estimated $\overline{x}$ = 3.957,SE = 0.218) reveals overall higher Eerie scores than Photorealistic Appearance (estimated $\overline{x}$ = 3.707, SE = 0.218).
In addition, a tendency for Animation Realism can be observed as well. Overall, Low Animation Realism led to higher Eerie scores (estimated $\overline{x}$ = 3.919, SE = 0.177) than Full Animation Realism (estimated $\overline{x}$ = 3.823, SE = 0.168). There was a noticeable peak in the *Happy* in the Semi-realistic Appearance condition acting against the trend however.
Additional two-way mixed ANOVA tests were conducted in a follow-up investigation to analyse the sample groups for the Emotional Scenarios separately - with a non-parametric equivalent test for the Emotion Scenarios *Neutral* and *Angry* as only here the assumption of homogeneity was violated.
While a two-way mixed ANOVA with between-subject factor Appearance Realism and within-subject factor Animation Realism revealed no significant overall effects for any of the Emotion Scenarios - for the Emotion Scenario *Happy* a planned pairwise comparisons test did reveal statistical significance for the interaction effect of Animation Realism with Appearance Realism. Specifically, a pairwise comparison with a Bonferroni adjusted 95% confidence interval highlighted that in the *Full* Animation Realism condition, the Appearance condition *Semi-realistic* was perceived to be significantly more eerie than the Appearance condition *Photorealistic* (p = 0.03). This becomes apparent in Figure 4 as well, where there is a peak visible for the Eerie measure in the sample group for Semi-realistic Appearance with *Full* Animation Realism for the *Happy* scenario ($\overline{x}$ = 4.58, SD = 1.48) when compared to its Photorealistic counterpart ($\overline{x}$ = 3.71, SD = 1.596).

4.3. Intensity (H3)

An additional point of investigation was to analyze in what capacity Animation Realism and Appearance Realism would impact perceived emotion intensity within different Emotion Scenario contexts.
Here, a three-way ANOVA didn't reveal any significant effects

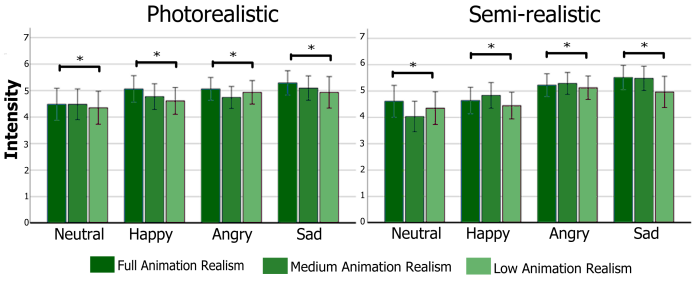


Fig. 5. Estimated marginal means for the variable Intensity.

for Animation Realism or Appearance Realism. With animation realism being rather close to significance in the ANOVA however (p = 0.08), a planned pairwise comparisons test was conducted with 95% Bonferroni adjusted confidence intervals. These highlighted a statistically significant effect of Animation Realism which wasn't readily apparent in the full three-way ANOVA. Specifically, the *Low* Animation Realism condition was overall rated significantly lower in intensity than the *Full* Animation Realism condition (p = 0.025). These significant effects become apparent in the visualization of Figure 5 as well. Particularly through the estimated means for the *Full* (estimated $\overline{x}$ = 4.988, SE = 0.127), *Medium* (estimated $\overline{x}$ = 4.843, SE = 0.133) and *Low* (estimated $\overline{x}$ = 4.718, SE = 0.144) Animation Realism conditions.
No statistically significant main effect for Appearance Realism was recorded.
Aside from the hypothesis investigations, there was also a statistically significant main effect of Emotion Scenario ($F(2.699, 161.97) = 11.756, p < 0.001$). Post hoc pairwise comparisons clarified that participants perceived the *Sad* Emotion Scenario to be significantly more intense than the *Neutral* ($p < 0.001$) and *Happy* ($p < 0.001$) Emotion Scenarios. Similarly, the *Angry* Emotion Scenario was perceived to be significantly more intense than the Emotion Scenarios *Neutral* ($p < 0.001$) and *Happy* (p = 0.013).
No interaction effects proved significant for the measure Intensity.

4.4. Realism Measures

To verify intended stimuli effects and account for possible confounding effects, additional realism measures were observed and analyzed.

Appearance Realism Perception (H4a)
There were no significant main effects for the independent variable Animation Realism.
Interestingly, there was no statistically significant main effect for the independent variable Appearance Realism as well. Figure 6 highlights that the estimated averages for the *Photorealistic* Appearance condition ($\overline{x}$ = 4.8, SE = 0.203) and the *Semi-realistic* Appearance condition ($\overline{x}$ = 4.478, SE = 0.203) do show a visual difference however.
In fact, there was an interaction effect of Animation Realism with Appearance Realism and Emotion Scenario which was on the cutoff for statistical significance

($F(6, 360) = 1.708, p = 0.052$). This result warranted further post hoc investigations. For that, pairwise comparisons were made, using a Bonferroni adjusted 95% confidence interval. Subsequently, this highlighted that indeed there were multiple significant interaction effects of Animation Realism with Appearance Realism for the measured variable Appearance Realism Perception as highlighted in Table 4.
Notably, separate two-way ANOVA tests for the Emotion Scenarios showed significant effects in post hoc pairwise comparisons with 95% Bonferroni adjusted confidence intervals. Results highlighted that for the *Neutral* Emotion Scenario, only in the Semi-realistic condition, the Full Animation Realism the character was perceived as having significantly higher appearance realism than the with Medium Animation Realism ($p = 0.003$). Additionally, for the *Angry* Emotion Scenario, only with Full Animation Realism was the Photorealistic character perceived significantly more realistic in appearance than the Semi-realistic character ($p = 0.037$).
Finally, it was important to verify that the *Photorealistic* Appearance Realism condition was perceived as realistic enough to warrant the *Photorealistic* classification. With an estimated mean clearly above average, the results do indicate that the photorealistic rendition of the virtual human was seen as very realistic - with room for improvement however.

Face Movement Realism (H4b)
The three-way ANOVA revealed no statistically significant main effects for Animation Realism or Appearance Realism.
Yet, a statistically significant interaction effect of Animation Realism and Emotion Scenario was revealed ($F(5.597, 335.841) = 2.234, p = 0.044$).
This warranted separate two-way ANOVA tests for all Emotion Scenarios which showed that for the *Sad* Emotion Scenario, a main effect of Animation Realism could be observed ($F(2, 120) = 10.197, p = 0.002$). Post hoc pairwise comparisons with Bonferroni 95% confidence intervals highlighted that the Face Movement Realism scores dropped significantly for the condition *Low* as compared to the scores for the *Medium* ($p = 0.024$) and *Full* ($p = 0.007$) Animation Realism conditions.
Interestingly, no other interaction effects proved statistically significant for the measure of Face Movement Realism.
Next to distinctly perceived Animation Realism levels, it was also important to verify that the *Full* Animation Realism condition was perceived as highly realistic. This gave an indication in how successful the performance-driven animation was in re-producing realistic facial motions. With an estimated $\overline{x} = 4.56$ ($SE = 0.162$), it becomes apparent that while the animation quality was overall deemed above average, its quality can be improved significantly as well.

Behavior realism (H4c)
A significant main effect of Animation Realism was found ($F(1.977, 118.62) = 3.23, p = 0.04$). Post hoc pairwise comparisons then showed that the *Low* Animation Realism condition led to significantly lower ratings for Behavior Realism than the conditions *Medium* ($p = 0.04$) and *Full* ($p = 0.015$).
No other main or interaction effects proved significant for the measure behavior realism.

Overall realism (H4d)
For the measure Overall Realism, there was a statistically significant main effect of Animation Realism ($F(1.889, 113.34) = 4.417, p = 0.014$). Subsequent post hoc pairwise comparisons highlighted that the *Low* Animation Realism condition led to significantly lower Overall Realism ratings than the *Full* Animation Realism condition ($p = 0.004$).
No other main or interaction effects proved statistically significant for the measure of Overall Realism.

Voice Realism and Body Movement Realism
In line with expectations, no statistically significant main or interaction effects could be observed for the measures Voice Realism and Body Movement Realism.

4.5. Qualitative Feedback

In addition to the reported quantitative results, participants had the optional opportunity to leave qualitative feedback as well. Most of it was simple positive feedback for the nature of the questionnaire as it was deemed ”interesting” and ”nice” by numerous participants. However, some feedback relayed further information. One of those notes indicated that the audio seemed to be perceived as slightly unsynced. This hints at the lip sync having room for improvement as it was noticeable enough for participants to mention it. Overall however, the survey was well received with some participants pointing out that they could tell there was a lot of care put into the creation of the emotion scenarios with the virtual human.

5. Discussion

Results highlighted that appearance realism had a significant effect on social presence. This effect was observable regardless of emotion scenario context as well, directly challenging past research insights which found that manipulating appearance realism wouldn't influence perceived social presence with a virtual character to a significant extent (see Zibrek et al. [16] [31]). Past studies investigating social presence were using characters more comparable in appearance realism to the semi-realistic character in this study. This hints at photorealism today to be so alike real humans that a significant increase in social presence has been reached.
Further, while there was no significant main effect found for animation realism, there was an interesting interaction effect between animation realism and appearance realism of the character. Specifically, when the animation was the most realistic the photorealistic character was inducing significantly higher perceptions of social presence than its semi-realistic counterpart. In conjunction with results from affinity and realism variables, this hints at a more interesting connection between appearance and animation realism which can be very helpful in designing virtual humans in the digital sphere.

Table 4. Statistical Significance - Part 1

Effect	ANOVA Result	Post hoc
Social Presence		
SOCIAL PRESENCE: Appearance Realism	$F(1, 60) = 4.057, p = 0.048$	Photorealistic appearance leads to higher social presence ($p = 0.048$).
SOCIAL PRESENCE: Appearance Realism x Animation Realism	$F(2, 120) = 1.55, p > 0.05$	Despite ANOVA being non-significant, pairwise Bonferroni 95% adjusted comparison showed: With *Full* Animation Realism *Photorealistic* characters lead to significantly higher social presence ($p = 0.038$).
SOCIAL PRESENCE [Separate test for *Angry*]: Appearance Realism x Animation Realism	$F(1.713, 102.806) = 4.637, p > 0.05$	Despite ANOVA being non-significant, pairwise Bonferroni 95% adjusted comparison showed: With *Medium* Animation Realism *Photorealistic* characters lead to significantly higher social presence ($p = 0.035$).
Affinity		
APPEAL: Emotion Scenario	$F(2.586, 155.16) = 10.976, p < 0.001$	Emotion Scenario *Angry* significantly less appealing than *Neutral* ($p < 0.001$) and *Happy* ($p < 0.001$). Emotion Scenario *Sad* also significantly less appealing *Neutral* ($p = 0.042$) and *Happy* ($p = 0.027$) scenarios.
EERIE [Separate test for *Happy*]: Appearance Realism x Animation Realism	$F(2, 120) = 0.728, p > 0.05$	Despite ANOVA being non-significant, pairwise Bonferroni 95% adjusted comparison showed: *Semi-realistic* character perceived to be significantly more eerie than *Photorealistic* character only in *Happy* scenario with *Full* Animation Realism ($p = 0.03$).
ATTRACTIVE: Animation Realism	$F(1.93, 115.8) = 3.527, p = 0.03$	*Low* Animation Realism significantly less attractive than *Medium* ($p = 0.025$) and *Full* Animation Realism ($p = 0.036$).
Intensity		
INTENSITY: Emotion Scenario	$F(2.699, 161.97) = 11.756, p < 0.001$	Emotion Scenario *Sad* perceived as significantly more intense *Neutral* ($p < 0.001$) and *Happy* ($p < 0.001$) scenarios. Emotion Scenario *Angry* perceived as significantly more intense than scenarios *Neutral* ($p < 0.001$) and *Happy* ($p = 0.013$).
INTENSITY: Animation Realism	$F(1.920, 115.219) = 3.225, p = 0.045$	*Full* Animation Realism significantly more intense than *Low* Animation Realism ($p = 0.025$).

With regards to appeal, while no statistical significance for animation realism could be found, the graphs in Figure 4 depict a clear tendency where higher animation realism led to higher appeal of the virtual character. Additionally, data revealed the photorealistic character to be appealing on average than the the semi-realistic version. Although not significantly so.

Further, the results for appeal are interestingly complemented by those for eeriness ratings as well. An interaction effect for emotion scenario, animation realism and appearance realism revealed that only when the animation realism was higher, the semi-realistic virtual human expressing happiness was perceived as significantly more eerie than its photorealistic counterpart. This peculiar result hints at a more complex explanation for eeriness perceptions. Specifically, the data suggests that photorealistic virtual humans are a much better fit for performance-driven highly realistic animations. These results supports past research insights by Bailenson et al. [32] who found that a mismatch between behavior realism and appearance realism can lead to lower social presence and lower affinity towards the virtual character. As animation realism in this study turned out to be an accurate predictor for perceived behavior realism, the comparison between studies becomes plausible.

With the addition of higher animation realism leading to significantly higher perceived attractiveness ratings, we can see that

Table 5. Statistical Significance - Part 2

Effect	ANOVA Result	Post-hoc
Realism		
APPEARANCE REALISM PERCEPTION: Appearance Realism x Animation Realism x Emotion Scenario	$F(6, 360) = 1.708$, $p = 0.052$	Emotion Scenario *Neutral* had overall significantly higher scores for Appearance Realism than the Emotion Scenario *Angry* ($p = 0.032$) only for *Semi-realistic* character with *Full* Animation Realism.
APPEARANCE REALISM PERCEPTION [Separate test for *Neutral*]: Appearance Realism x Animation Realism	$F(1.75, 105.075) = 2.93$, $p = 0.06$	*Full* Animation Realism condition led to higher perceived appearance realism than *Medium* Animation Realism ($p = 0.003$) only for *Semi-realistic* character.
APPEARANCE REALISM PERCEPTION [Separate test for *Angry*]: Appearance Realism x Animation Realism	$F(2, 120) = 2.791$, $p = 0.06$	*Photorealistic* character elicited higher perception of appearance realism than *Semi-realistic* character ($p = 0.037$) only with *Full* Animation Realism.
FACE MOVEMENT REALISM [Separate test for *Sad*]: Animation Realism	$F(2, 120) = 10.197$, $p = 0.002$	*Low* Animation Realism perceived to be significantly less realistic than *Medium* ($p = 0.024$) and *Full* ($p = 0.007$) Animation Realism.
BEHAVIOR REALISM: Animation Realism	$F(1.977, 118.62) = 3.23$, $p = 0.04$	*Low* Animation Realism led to significantly lower perceived behavior realism than *Medium* ($p = 0.04$) and *Full* ($p = 0.015$)
OVERALL REALISM: Animation Realism	$F(1.889, 113.34) = 4.417$, $p = 0.014$	*Low* Animation Realism led to significantly lower perceived overall realism than *Full* ($p = 0.004$)

animation realism plays an important role in conjunction with appearance realism. Eeriness ratings were only significantly increased in one case for the semi-realistic virtual human when a mismatch between appearance realism and animation realism was apparent.

Results regarding eeriness perceptions differed from the study by Tinwell et a.[8]. where significance for animation realism was found. It is to be noted however that the stimuli participants were subject to differed between the two studies. In Tinwell's study, upper face motion was an important indicator the the virtual agent's reaction to external stimuli. The stimuli simulated unexpected sounds which would require surprise reactions from the virtual agents. Environment scenario context beyond the virtual human's emotional state might have an interaction effect with animation realism as well - this was not investigated and outside the scope of this study.

However, it is an important insight to realize that animation realism shouldn't be limited to lip sync only. While the removal of eyebrow motion didn't lead to significantly different eerie ratings, there was a significant change in attractiveness perceptions. Furthermore, manipulating animation realism led to a significant main effect for perceived differences in behavior realism and overall realism. Therefore, more realistic motions are perceived as more attractive which supports studies by Zibrek et al. [17] [33]. As attractiveness has been linked to more positive customer engagement [34], it can be a desired trait for embodied virtual assistants.

Finally, the emotion intensity was perceived differently based on emotion scenario. This result needs to be regarded critically however. While it is true that the angry and sad emotion scenarios were perceived to be a lot more intense - this might simply be due to the scenario script resonating more with participants or the voice acting leading to these conclusions. The observation that is more poignant is that the emotions were perceived as significantly less intense for the *Low* animation realism condition.

We would expect upper face motion to be incredibly important in portraying and displaying emotion. Complete removal of upper face motion did indeed lead to significantly lower intensity ratings. Interestingly enough however, the lack of eyebrow motion didn't lead to significantly different perceptions of emotion intensity. While the averages for the different animation realism conditions do show a difference between all three levels of animation realism, the influence of eyebrow motion didn't lead to differences that were big enough to warrant statistical significance.

This might be due to the short length of the video stimuli however. As the scenes were limited to around 20 seconds, this

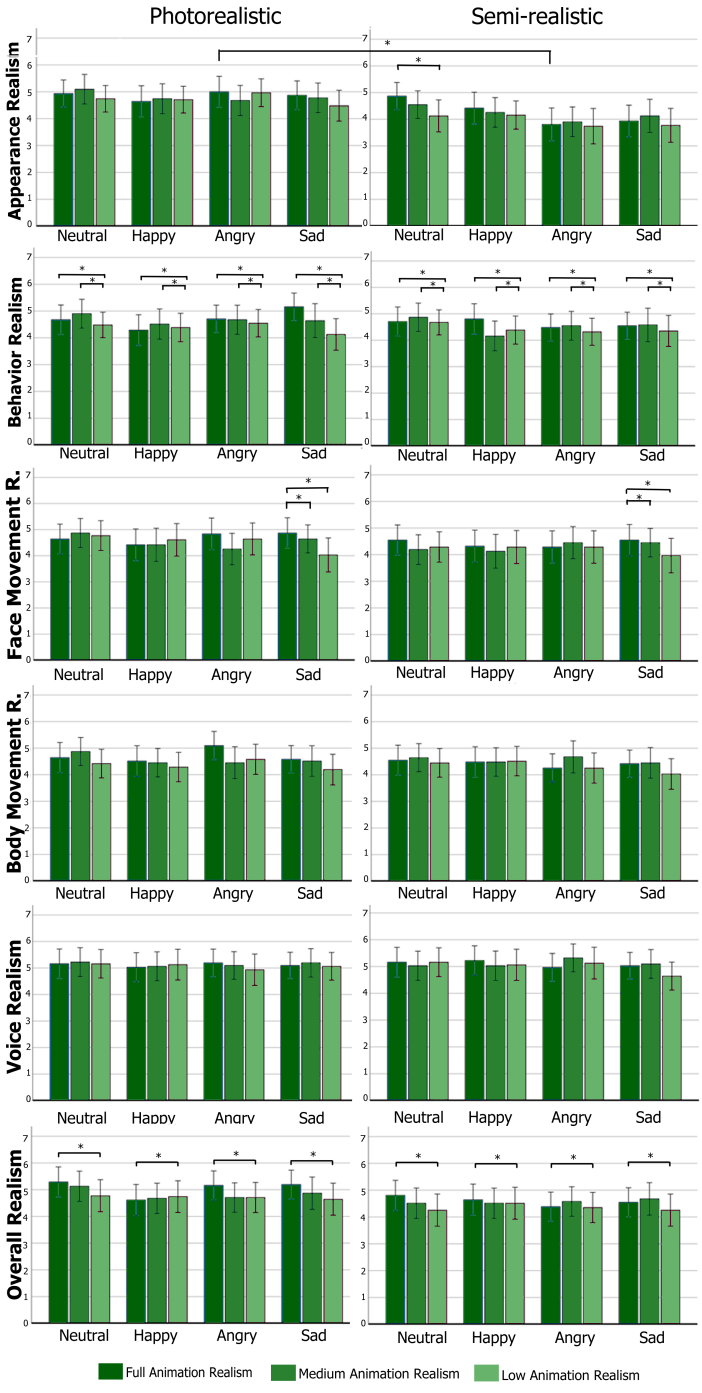


Fig. 6. Estimated marginal means for the realism check variables.

observation might be specific to short visual stimuli only. Additionally, the eyebrow motion of humans is quite subtle which was reflected in the video renders of this study. The addition or removal of upper-face wrinkles was quite pronounced between the conditions with and without eyebrow motion however. This hints towards upper face wrinkles not being relevant in short animation sequences with regards to emotion intensity.
Finally, the measure of Face Movement Realism revealed that manipulating animation realism only led to signficant differences in the *Sad* emotion scenario. This might be due to the body movement of the virtual human varying between emotion scenarios however. The sad scenario allowed for the most subdued body motions, giving participants the opportunity to focus more on the virtual agent's face.

6. Limitations and future work

These results need to be put into perspective however. The experiment was conducted with only one female virtual human model. It can be interesting to investigate if and how people would react differently to other virtual humans with regards to ethnic background and gender. Ethnic biases might have played a role in how the virtual human was perceived. Varying facial morphology to reflect ethnic differences can lead to further insight into the influence of ethnic bias on the perception of virtual characters.
This work also leads to the question of whether face-tracking and retargeting systems might benefit from creating altered algorithms for different ethnic groups for better animation quality as has been proposed in the past [35] as challenges arose from using a dark-skinned actress.
Further, this study was limited to an on-screen experiment due to the highest quality level of detail for Epic Games' MetaHumans not running reliably enough within real-time VR applications at the time of publication. Researching the interaction effects of animation realism and appearance realism in a VR environment might be a valuable endeavor.
Moreover, in this study, the background of the virtual human was limited to a blank screen. It can also be an interesting research question to investigate the effect an environment has on the perception of an emotionally expressive virtual human.
Additionally, this research didn't allow for a responsive virtual human which could engage and react to participants. It was limited to an on-screen video presentation. This gives rise to the question how the realism perception of virtual humans is influenced in an interactive setting.
In closing, there are direct criticisms to this study that need to be addressed as well. The lip sync was not perfect and had room for improvement, as indicated by participant feedback. Lip sync overall leads to the question in how far the effect of voice is prevalent in the perception of virtual humans. While there has been evidence for a voice - appearance mismatch to lead to different perceptions towards the virtual human, this was outside of the scope for this study and has not been investigated. As our perception is naturally cross-modal [36], voice has a definite effect on perception [15][37]. A follow-up study could investigate in how far simultaneous manipulation of voice realism has an influence on the perception of the virtual human as well. Though it is to be noted that a survey as used in this study might not be sufficient to capture cross-modal dependencies. This is due low-level perceptual dependencies in the human brain with neurons that activate for cross-modal audio-visual stimulation [38].
Moreover, in this study only one way of manipulating appearance realism was accounted for. Only by varying levels of detail to influence photorealism level was the influence of appearance investigated. There are further comparisons to be made, for

example comparing the photorealistic version with a stylized, more cartoonish version or with anthropomorphic characters as well.
Additionally, animation realism in this study was manipulated in a specific way - manipulations with more granularity and more variation might lead to new or different insights.
To be noted is that for a number of tests during the statistical analysis, the assumption of homogeneity for ANOVA had been violated. This led to the use of a non-parametric solution which does not have as much power as a robust ANOVA. Therefore, a repeat study under similar conditions with sample groups that conform to the homogeneity of variance assumption is recommended. This could lead to further insights and provide more data towards this research topic.
Finally, the current state of technology opens up very relevant and interesting ethical questions with regards to digital humans. We are at a stage where virtual humans are realistic enough to replace real actors in movies and TV [39] without the audience being the wiser. It is difficult to predict the consequences on the artist's workflow or on the impact for ownership over digital assets. Selling your digital human copy on the market is on the precipice of becoming reality [40].
Now that virtual humans are almost indistinguishable from real humans, evidenced by significantly increased social presence compared to what technology was capable of before - how ethical is it to create digital humans in the likeness of real people to use in digital media?

7. Conclusion

In this study, the main and interaction effects of a virtual human's appearance realism and animation realism within different emotion scenario contexts were investigated.
The study reveals interesting interaction effects between appearance realism and animation realism for emotionally expressive virtual humans. While only appearance realism lead to a significant main effect on perceived social presence, this study further highlighted important conjuntion effects of animation realism with appearance realism for social presence.
Higher movement realism has been found to be significantly more attractive. However, with regards to the uncanny valley, this study highlighted a more complex interaction effect between animation realism and appearance realism for perceived eeriness which warrants further investigation. Only with full animation realism was the semi-realistic character perceived to be significantly more eerie than the photorealistic character.
Together with results regarding social presence, it suggests that higher motion realism is only effective when matched with higher appearance realism.
Finally, evidence in this study also suggests that we prefer happy or neutral emotion scenarios for virtual humans. For emotional support chatbots for example, social presence is incredibly important in facilitating better treatment efficacy [41]. Furthermore, results here suggest that for shorter animation sequences, removing eyebrow motion will not lead to significantly different perceived emotion intensity. This might inform animators and character artists that for short sequences, the intended portrayed emotions would be sufficiently conveyed without eyebrow motion as well.

References

[1] Amazon Inc., . Echo & Alexa Devices. (n.d.). URL: https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b?ie=UTF8&node=9818047011; last accessed 25 September 2022.

[2] Google, . Google Assistant. (n.d.). URL: https://assistant.google.com/; last accessed 25 September 2022.

[3] Epic Games, . MetaHuman Creator. 2021. URL: https://www.epicgames.com/.

[4] Ziva Dynamics, . Ziva Face Trainer. 2022. URL: https://zivadynamics.com/ziva-face-trainer.

[5] REBLIKA, . Valentina 4D. 2020. URL: https://reblika.com/projects/valentina-4d/.

[6] Ferstl, Y, Thomas, S, Guiard, C, Ennis, C, McDonnell, R. Human or robot? investigating voice, appearance and gesture motion realism of conversational social agents. In: Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents. IVA '21; New York, NY, USA: Association for Computing Machinery. ISBN 9781450386197; 2021, p. 76–83. URL: https://doi.org/10.1145/3472306.3478338. doi:10.1145/3472306.3478338.

[7] Higgins, D, Egan, D, Fribourg, R, Cowan, B, McDonnell, R. Ascending from the valley: Can state-of-the-art photorealism avoid the uncanny? In: ACM Symposium on Applied Perception 2021. SAP '21; New York, NY, USA: Association for Computing Machinery. ISBN 9781450386630; 2021,URL: https://doi.org/10.1145/3474451.3476242. doi:10.1145/3474451.3476242.

[8] Tinwell, A, Nabi, DA, Charlton, JP. Perception of psychopathy and the uncanny valley in virtual characters. Computers in Human Behavior 2013;29(4):1617–1625. URL: https://www.sciencedirect.com/science/article/pii/S0747563213000113. doi:https://doi.org/10.1016/j.chb.2013.01.008.

[9] Seymour, M, Yuan, L, Dennis, A, Riemer, K. Face it, users don't care: Affinity and trustworthiness of imperfect digital humans. In: Proceedings of the 55th Hawaii International Conference on System Sciences. 2022,.

[10] McDonnell, R, Breidt, M, Bülthoff, HH. Render me real? investigating the effect of render style on the perception of animated virtual humans. ACM Trans Graph 2012;31(4). URL: https://doi.org/10.1145/2185520.2185587. doi:10.1145/2185520.2185587.

[11] Zibrek, K, Martin, S, McDonnell, R. Is photorealism important for perception of expressive virtual humans in virtual reality? ACM Trans Appl Percept 2019;16(3). URL: https://doi.org/10.1145/3349609. doi:10.1145/3349609.

[12] Mori, M. The uncanny valley: the original essay by masahiro mori. IEEE Spectrum 1970;.

[13] MacDorman, KF, Green, RD, Ho, CC, Koch, CT. Too real for comfort? uncanny responses to computer generated faces. Computers in human behavior 2009;25(3):695–710.

[14] Seyama, J, Nagayama, RS. The uncanny valley: Effect of realism on the impression of artificial human faces. Presence 2007;16(4):337–351.

[15] Higgins, D, Zibrek, K, Cabral, J, Egan, D, McDonnell, R. Sympathy for the digital: Influence of synthetic voice on affinity, social presence and empathy for photorealistic virtual humans. Computers Graphics 2022;104:116–128. URL: https://www.sciencedirect.com/science/article/pii/S0097849322000474. doi:https://doi.org/10.1016/j.cag.2022.03.009.

[16] Zibrek, K, McDonnell, R. Social presence and place illusion are affected by photorealism in embodied vr. In: Motion, Interaction and Games. MIG '19; New York, NY, USA: Association for Computing Machinery. ISBN 9781450369947; 2019,URL: https://doi.org/10.1145/3359566.3360064. doi:10.1145/3359566.3360064.

[17] Zibrek, K, Niay, B, Olivier, AH, Hoyet, L, Pettre, J, McDonnell, R. The effect of gender and attractiveness of motion on proximity in virtual reality. ACM Transactions on Applied Perception (TAP) 2020;17(4):1–15.

[18] Biocca, F. The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments [1]. Journal of Computer-Mediated Communication 1997;3(2). URL: https://doi.org/10.1111/j.1083-6101.1997.tb00070.x. doi:10.1111/j.1083-6101.1997.tb00070.x; jCMC324.

[19] Biocca, F, Harms, C, Burgoon, JK. Toward a more robust theory and measure of social presence: Review and suggested criteria. Presence: Teleoperators & virtual environments 2003;12(5):456–480.

[20] Guimarães, M, Prada, R, Santos, PA, Dias, Ja, Jhala, A, Mascarenhas, S. The impact of virtual reality in the social presence of a virtual agent. In: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents. IVA '20; New York, NY, USA: Association for Computing Machinery. ISBN 9781450375863; 2020,URL: https://doi.org/10.1145/3383652.3423879. doi:10.1145/3383652.3423879.

[21] Wisessing, P, Zibrek, K, Cunningham, DW, Dingliana, J, McDonnell, R. Enlighten me: Importance of brightness and shadow for character emotion and appeal. ACM Transactions on Graphics (TOG) 2020;39(3):1–12.

[22] Ben-David, BM, van Lieshout, PHHM, Leszcz, T. A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. Brain Injury 2011;25(2):206–220. URL: https://doi.org/10.3109/02699052.2010.536197. doi:10.3109/02699052.2010.536197. arXiv:https://doi.org/10.3109/02699052.2010.536197; pMID: 21117915.

[23] Dynamixyz, . Performer2SV. (n.d.). URL: https://www.dynamixyz.com/software2; version: v2.11.0.0.

[24] Autodesk, INC., . Maya. (n.d.). URL: https://autodesk.com/maya; version 2022.

[25] Qualtrics, . Qualtrics XM Platform. 2005. URL: https://www.qualtrics.com; version: August 2022.

[26] Bailenson, JN, Blascovich, J, Beall, AC, Loomis, JM. Interpersonal distance in immersive virtual environments. Personality and social psychology bulletin 2003;29(7):819–833.

[27] Sterna, R, Zibrek, K. Psychology in virtual reality: toward a validated measure of social presence. Frontiers in Psychology 2021;12.

[28] Buhrmester, M, Kwang, T, Gosling, SD. Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? Perspectives on Psychological Science 2011;6(1):3–5. URL: https://doi.org/10.1177/1745691610393980. doi:10.1177/1745691610393980. arXiv:https://doi.org/10.1177/1745691610393980; pMID: 26162106.

[29] Noguchi, K, Gel, YR, Brunner, E, Konietschke, F. nparLD: An R software package for the nonparametric analysis of longitudinal data in factorial experiments. Journal of Statistical Software 2012;50(12):1–23. URL: https://www.jstatsoft.org/v50/i12/.

[30] Konietschke, F, Placzek, M, Schaarschmidt, F, Hothorn, LA. nparcomp: An R software package for nonparametric multiple comparisons and simultaneous confidence intervals. Journal of Statistical Software 2015;64(9):1–17. URL: http://www.jstatsoft.org/v64/i09/.

[31] Zibrek, K, Kokkinara, E, McDonnell, R. Don't stand so close to me: investigating the effect of control on the appeal of virtual humans using immersion and a proximity-based behavioral task. In: Proceedings of the ACM Symposium on Applied Perception. 2017, p. 1–11.

[32] Bailenson, JN, Swinth, K, Hoyt, C, Persky, S, Dimov, A, Blascovich, J. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments 2005;14(4):379–393. URL: https://doi.org/10.1162/105474605774785235. doi:10.1162/105474605774785235.

[33] Zibrek, K, Niay, B, Olivier, AH, Pettré, J, Hoyet, L, Mcdonnell, R. Proximity in vr: The importance of character attractiveness and participant gender. In: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE; 2022, p. 672–673.

[34] Fang, S, Zhang, C, Li, Y. Physical attractiveness of service employees and customer engagement in tourism industry. Annals of Tourism Research 2020;80:102756. URL: https://www.sciencedirect.com/science/article/pii/S0160738319301136. doi:https://doi.org/10.1016/j.annals.2019.102756.

[35] Salah, AA, Alyüz, N, Akarun, L. Registration of three-dimensional face scans with average face models. Journal of Electronic Imaging 2008;17(1):011006.

[36] Stein, BE, Meredith, MA. The merging of the senses. The MIT press; 1993.

[37] Brambilla, M, Masi, M, Mattavelli, S, Biella, M. Faces and sounds becoming one: Cross-modal integration of facial and auditory cues in judging trustworthiness. Social Cognition 2021;39(3):315–327.

[38] Shimojo, S, Shams, L. Sensory modalities are not separate modalities: plasticity and interactions. Current opinion in neurobiology 2001;11(4):505–509.

[39] Stuart, SC. 'the congress' offers robin wright the digital fountain of youth 2014;URL: https://www.pcmag.com/news/the-congress-offers-robin-wright-the-digital-fountain-of-youth.

[40] Santos, R. You can sell your face to companies who create deepfake marketing clones 2022;URL: https://thedebrief.org/sell-your-face-deepfake/.

[41] Tsai, WHS, Liu, Y, Chuan, CH. How chatbots' social presence communication enhances consumer engagement: the mediating role of parasocial interaction and dialogue. Journal of Research in Interactive Marketing 2021;.

Supplementary Material

Supplementary material is added in the form of video renders with the virtual human as presented to participants. It can be found here: https://drive.google.com/file/d/1lSnOXajMnlqhnyYcNkmrut08akSDPkLX/view?usp=sharing

Appendix A. Emotion Scenario Scripts

Appendix A.1. Neutral

"Hmm, so, I went to the grocery store this morning and I made sure to buy onions, some potatoes.. Umm.. Right. I also didn't forget about the bread. I bought some pasta as well.."

Appendix A.2. Happy

"Oh hi! It's so nice to see you. In general, today feels like a really good day. Umm, first the weather is really, really wonderful today. And.. The food I ate earlier at that restaurant I went to also tasted really nice. But most importantly, I got a promotion at work!"

Appendix A.3. Angry

"Alright, I'm done. You – Stop what you're doing and come over here. Listen. I am sick and tired of you being late every single time. We're supposed to be a team – but we can not count on you at all. I'm just done."

Appendix A.4. Sad

"Oh. Hey. It's just – I've been really lonely lately. Every day is the same.. I wake up. Sit down at my PC all day – go back to sleep. Repeat. It's just.."

Appendix B. Full statistical analysis results

For a number of sample groups for some dependent variables, the assumptions for ANOVA were violated. The assumption for homogeneity being violated meant having to use a non-parametric mixed ANOVA. In addition, violating the assumption of sphericity meant having to adjust degrees of freedom to Greenhouse Geisser or Huynh-Feldt estimates. They are highlighted in the following to further explain results as shown in the paper.

Appendix B.1. Social presence

Mauchly's test revealed that the assumption of sphericity had been violated for *Emotion Scenario* ($\chi^2 = 31.5, p < 0.001$), so degrees of freedom were adjusted using Huynh-Feldt estimates ($\epsilon = 0.786$).

Appendix B.2. Appeal

Levene's test revealed that for multiple sample groups, the homogeneity of variance assumption was violated. Therefore, an alternative non-parametric equivalent to ANOVA was used. This revealed a significant effect of Emotion Scenario ($F(2.586, 155.16) = 10.976, p < 0.001$). Subsequent post hoc investigations highlighted that the Emotion Scenarios *Angry* and *Sad* scored significantly lower on the Appeal scale than the Emotion Scenarios *Neutral* and *Happy* (see Table 4.

Appendix B.3. Attractive

For the measure Attractive, the assumption of homogeneity of variance was violated for one sample group (p = 0.028) as Levene's test revealed. Therefore, a non-parametric equivalent for a three-way mixed ANOVA was employed.

Appendix B.4. Eerie

As a Levene's test revealed that for the sample groups in regards to the Emotion Scenarios *Neutral* and *Angry* the assumption of homogeneity of variances was violated, a non-parametric three-way mixed ANOVA equivalent was run for these.

Appendix B.5. Intensity

Levene's test proved that for one sample group, the homogeneity of variance assumption was violated. A non-parametric equivalent to ANOVA was then employed which revealed a statistically significant main effect of Emotion Scenario ($F(2.699, 161.97) = 11.756, p < 0.001$). In addition, non-parametric pairwise comparisons highlighted a statistically significant effect of Animation Realism.

Appendix B.6. Face Movement Realism

The assumption of sphericity had been violated for Animation Realism ($\chi^2 = 7.032$, p = 0.03) and for the interaction effect of Emotion Scenario and Animation Realism ($\chi^2 = 35.36$, $p = 0.018$) as the Mauchly's test has shown. Degrees of freedom were adjusted with Huyhn-Feldt estimates ($\epsilon = 0.786$).

Appendix B.7. Behavior realism

For the measure Behavior Realism, Levene's test highlighted one sample group that violated the homogeneity of variances assumption ($p = 0.005$). Therefore, a non-parametric counterpart to the three-way mixed ANOVA was employed.

Appendix B.8. Overall realism

Levene's test revealed that one sample group violated the homogeneity of variances assumption ($p = 0019$). A non-parametric equivalent to the three-way mixed ANOVA was then used. This operation then revealed a statistically significant main effect of Animation Realism ($F(1.889, 113.34) = 4.417, p = 0.014$).

Appendix C. Performance-driven animation Workflow

Appendix C.1. Performance-driven facial animation

As the focus of this research was the analyzing of the perception of photorealistic virtual humans, the goal was to achieve highly realistic animations as well. Therefore, the study made use of performance-driven facial animation.
As for the facial animation, the Utrecht University Motion Capture Lab was used in order to capture the face motion and produce animations derived from that.

Appendix C.1.1. Setup

A Dynamixyz hardware and software setup was used to create facial animations. Specifically, a head-mounted camera (HMC) with Dynamixyz's Grabber software facilitated the wireless capture of face movements. It is to be noted that, the Dynamixyz face tracking system is markerless. This means that no reflective markers were placed on the actress' face. In addition, Performer2SV was used to retarget facial movements from the actress' face onto the MetaHuman.

Appendix C.1.2. Calibration

The HMC first needed to be calibrated with a checkerboard as is standard to adjust lens settings (intrinsic calibration). Ideally, the actor's face will remain in the exact same position relative to the HMC between capture sessions. This was concluded after several tests with the system. While it seems somewhat robust to small angle changes between captures, it was noticeable that specifically for lip syncing accuracy and eyebrow tracking, the system performed worse.

Appendix C.1.3. Tracking

In addition to the camera itself being calibrated, the system needs to calibrate to the actor's face as well. Salient feature points differ from face to face – some people's noses are bigger. Face shapes change from person to person. In addition, others have thinner eyebrows, thinner or thicker lips, or more pronounced cheekbones for example. The system needs to first learn where important feature points are in the actor's face.
For that, select frames are chosen from so-called "Range-Of-Motion" (ROM) video captures made with the HMC. The goal is to have a wide variety of facial expressions for the system to learn from all the possible configurations of one's face. Teaching the system all of the movements possible with one's face through isolated frames, it enables the setup process to correctly track feature points over video lengths as well. Using these feature points, the Dynamixyz system internally creates a 3D model matching the actor's face.
The ROM in this study to teach the system the actress' feature points was made up from Dynamixyz ROM guidelines as well as MetaHumans' facial ROM as shown on the MetaHuman Creator website. An additional ROM video of such has been added to the supplementary material for demonstration purposes.

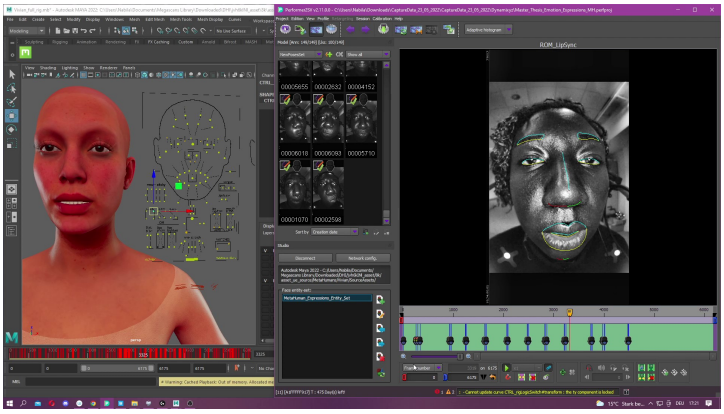


Fig. C.7. Screenshot of a retargeting session from Performer2SV to Maya

Appendix C.2. Retargeting

The term "retargeting" describes the process of translating motions from one joint or skeleton structure to another. In this case, the system had to translate from the actress' face onto a MetaHuman.
For this process, Dynamixyz requires a control rig for the virtual character. With MetaHumans being so complex, including 669 blendshapes as well as 157 controls, keep in mind that the retargeting process is very time consuming (see Figure C.7).

- In a very first step, pre-requisites need to be installed for Autodesk Maya and Quixel Bridge – the software needed to export MetaHumans. Dynamixyz offered a plugin that is to be installed and activated in Maya for the retargeting process.
- Then, the MetaHuman needs to be loaded into Autodesk Maya. This is achieved through exporting the virtual human from the software Quixel Bridge onto Maya. Quixel Bridge is required to install a Maya plugin for a successful export of the character.
- To connect the Dynamixyz software Performer2SV to Maya, the Dynamxyz plugin needs to be activated by inputting "dxyzBridge" into the MEL command line of Maya. It opens an application which can establish a socket connection between Performer2SV and Maya.
- The Dynamixyz system has three solvers to match movements from the actor to the virtual character and produce fitting facial animations. One solver for the eye-gaze animation, one solver for the upper face animation (everything above the nose excluding eye-gaze) and one solver for the lower face animation. In this second step, Dynamixyz needs to know which controls are to be assigned to which solvers. This assignment must be done manually.
- After assigning the controllers to the proper solvers, the retargeting menu in Performer2SV becomes available. Here, the same frames as chosen for the tracking setup are displayed. Using the controllers in Maya, you now manipulate the MetaHuman's face to match the actor's expression in Performer2SV as shown in Figure C.7. Through the socket connection, Performer2SV can read the live input happening within the Maya software.
- You can manually select whether the Dynamixyz software uses the frame and corresponding MetaHuman control rig state to learn for the upper face solver, the lower face solver or the gaze solver. Any combination of the three solvers is possible – meaning that if a frame is specifically for lip sync purposes only, you can select for the system to use the frame only for the lower face solver for example. In another example, it is also possible to select for the system to use the frame for upper face and eye-gaze solvers.
- To ensure that the system doesn't learn any unwanted correlation between upper face, lower face and eye-gaze states, it was found through experimentation that there should be a sufficient amount of retargeting links that only focus on lower face, upper face and eye-gaze solvers in isolation as well.
- In total, around 250 retargeting links resulted from the ROM sequences to ensure a base-line for appropriate lip syncing and proper translation from complex expressions onto the MetaHuman.

Appendix C.3. Choosing a MetaHuman

The MetaHuman as chosen for this study was used due to the following reasons:

- To facilitate a repeatable study, an officially pre-configured released MetaHuman by Epic Games was chosen, named Kioko.
- She was the latest released MetaHuman at the time of the experiment, specifically for Unreal Engine 5. As the characters released only for Unreal Engine 5 have higher texture detail, they seem more realistic to the human eye.
- Due to Kioko having a similar facial structure as the actress in the study, the retargeting process would be easier.

Appendix C.3.1. Capture

The workflow for the performance-driven animation after completing the setup and registering retargeting links for ROM sequences, then consists of recording video of emotion scenario using a prepared script and adjusting lip sync retargeting links to better fit the video content.

Appendix C.4. Body Motion Capture

Along with performance-driven facial animation, the body motion was captured as well. This facilitated realistic upper body movement, even if the participants only saw the virtual human from the shoulders up.
For the body motion as well, the Utrecht University Motion Capture Lab was used in order to capture the body motion and produce animations derived from that.

Appendix C.4.1. Setup

For the body motion, a Vicon motion capture setup with 14 cameras was utilized. On the body suit, there were 74 reflective markers, arranged in the Vicon "Front-Waist-10-Fingers" setup which describes the production marker placement with additional markers for all 10 fingers.
For the study however, head markers were placed on a Dynamixyz HMC.

Appendix C.4.2. Calibrating

Before any movement can be captured, the cameras as well as the actor need to be calibrated. For one, the cameras need to understand their positioning in the 3D space (extrinsic calibration) and the capture space that they span. Further, every human build is different, therefore the system needs to learn where the actor's joints are located. The software used to track the reflective markers on the suit was Vicon Shogun Live. The procedure as described in [tutorial reference] was followed to ensure proper calibration.

Appendix C.4.3. Retargeting

The calibration steps would be sufficient to ensure capture data for Vicon skeletons – however, in this study MetaHumans were used. The skeleton structure of MetaHumans is different from Vicon skeletons. Therefore, proper a retargeting setup had to follow. The process of translating motion from one skeleton to another is referred to as 'retargeting'. The software Vicon Shogun Post was used for this process. Using Vicon Shogun Post, the retargeting process was as follows:

- Export MetaHuman skeleton using Maya. MetaHumans are limited by Epic Games to only be properly interactable within Unreal Engine and Autodesk Maya. An export in the proper format for Vicon Shogun Post is therefore most easily achieved by loading any MetaHuman into Maya. Within Maya, the bones and drivers as used for skinning should all be removed. (Their build is irrelevant, it's the bone names that are imperative.) This facilitates real-time retargeting in the Vicon software as the number of bones in MetaHuman skeletons is too high otherwise.
- Any recording done in Vicon Shogun Live that produces an mcp-file can be loaded into Vicon Shogun Post. The retargeting setup can then be started by uploading the MetaHuman skeleton.

Appendix D. Questionnaire

The questionnaire began with an ethics statement and consent form, guaranteeing that participants' data is deleted after the end of the study. Participants were then be asked the following questions regarding demographics information.
For every video, participants were to then answer attention check questions as well. These would change for ever emotion condition and animation realism condition. In the following a printout for the "Full Animation Realism" condition:

Introduction & Consent Form

Welcome to this study on 3D virtual humans as part of a master thesis research project conducted by Game and Media Technology Masters student Nabila Amadou, supervised by Dr. Zerrin Yumak at Utrecht University, The Netherlands.

In this study, you will be watching videos of a virtual human talking and after each video, there will be a couple of questions to answer regarding your perceptions of the virtual human's appearance and movement. At the beginning you will also be asked some demographic questions (gender, age and familiarity with virtual human technology). The whole study takes around 25 minutes.

Data collected in this study will be used only for scientific purposes. Your data will be stored on the Qualtrics server and will be saved in a secure local environment for further analysis. Your data will remain non-identifiable and it will be deleted after the end of the project. Anonymous data from this study may be shared in a public repository for research purposes and be presented in scientific publications.

For further questions, please contact n.amadou@students.uu.nl or

z.yumak@uu.nl.

Consent Form

Please read the statements below:

- I confirm that I have read and understood the information provided to me for this study.
- I understand that my participation is voluntary and I can withdraw at any time.
- I agree that research data gathered in this study may be published and may be shared in public repositories provided that my identity remains anonymous.

By clicking the "I Consent" button, you give your explicit consent to participate in this study. If you do not agree, you can simply leave the page.

◯ I Consent

Block 27

Please note that you will be watching 12 short videos.
There will be no duplicate videos as slight changes have been made on each one of them. Some of them will be very similar to each other in content, but please make sure to watch every single video to reflect your differing perceptions accurately in your responses.

At the end, you will be shown your unique participation ID to input as the survey code on Amazon Mechanical Turk.

Finally, thank you again for participating!

◯ I Understand

Demographic Information

What is your age?

What is your gender?

◯ Male
◯ Female
◯ Non-binary / third gender
◯ Prefer not to say

Do you have prior experience with virtual humans through films and/or games?

◯ None at all
◯ A little

- ◯ A moderate amount
- ◯ A lot
- ◯ A great deal

Photorealistic Appearance - Condition Neutral - Full Anim

Please watch the following video:

What was the topic the virtual human was talking about?

- ◯ The weather
- ◯ Groceries
- ◯ Looking for keys

Which emotion did you perceive in the virtual human?

- ◯ Angry

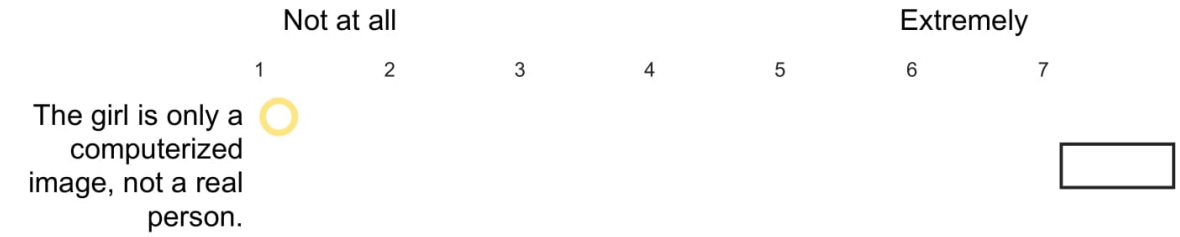
- ◯ Happy
- ◯ Sad
- ◯ Neutral
- ◯ Other

How intense was the emotion that you observed?

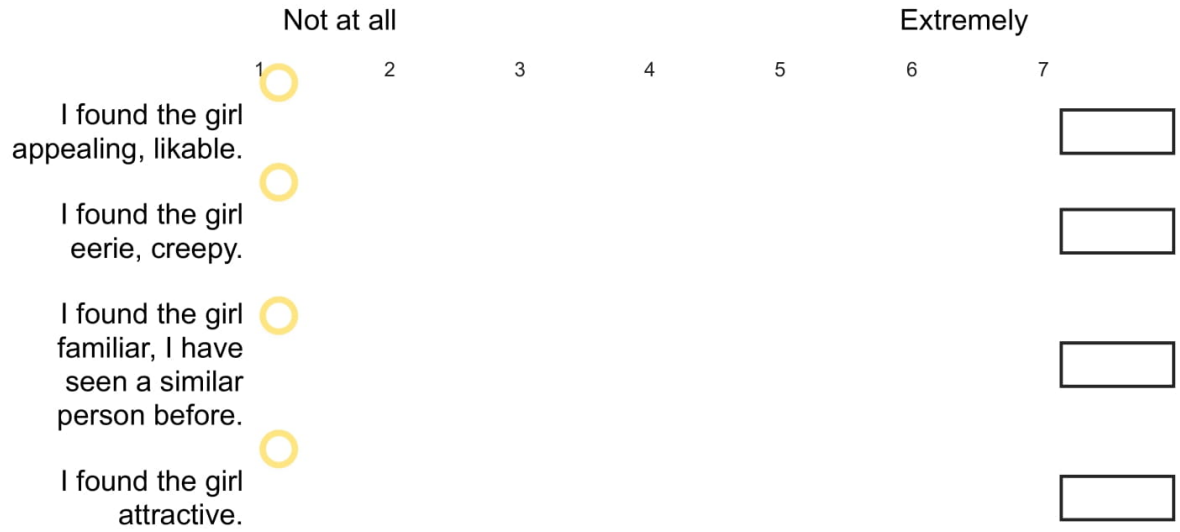
	Not at all						Extremely
	1	2	3	4	5	6	7
Intensity	◯	◯	◯	◯	◯	◯	◯

How much does each statement apply for you?

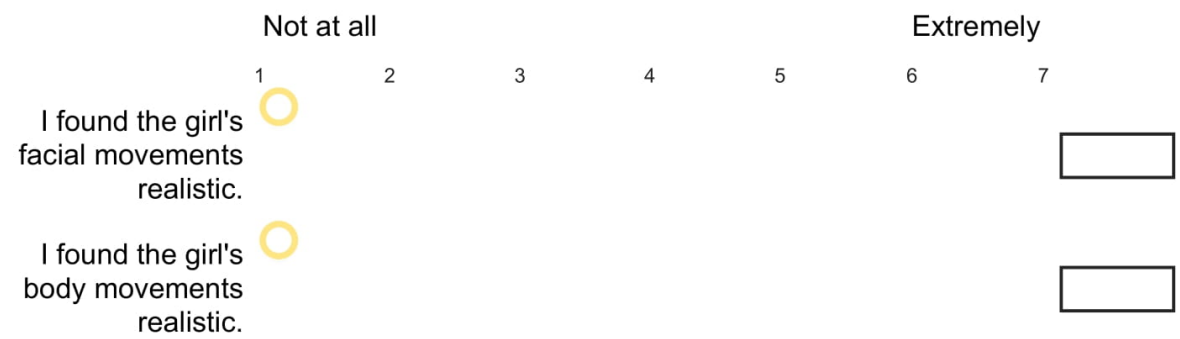
	Not at all						Extremely
	1	2	3	4	5	6	7
It feels as if I am in the presence of another person in the room with me.	◯	◯	◯	◯	◯	◯	◯
It feels as if the girl is watching me and is aware of my presence.	◯	◯	◯	◯	◯	◯	◯
The thought that the girl isn't real crossed my mind often.	◯	◯	◯	◯	◯	◯	◯
The girl appears to be alive.	◯	◯	◯	◯	◯	◯	◯

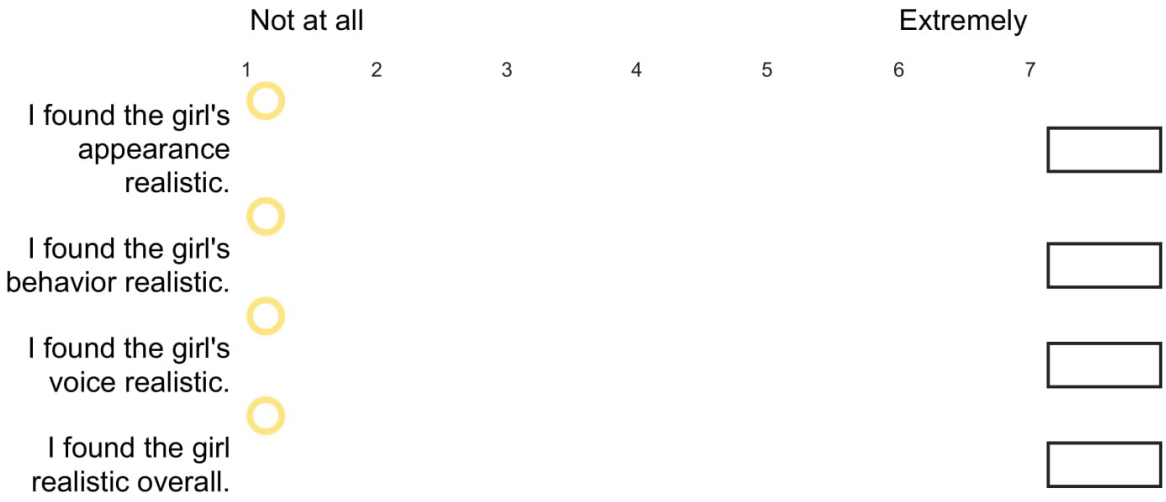


How much does each statement apply for you?



How much does each statement apply for you?





Photorealistic Appearance - Condition Happy - Full Anim

Please watch the following video:

What topic did the virtual human mention?

- ◯ A work promotion
- ◯ Creativity
- ◯ A new pet

Which emotion did you perceive in the virtual human?

- ◯ Angry
- ◯ Happy
- ◯ Sad
- ◯ Neutral
- ◯ Other

How intense was the emotion that you observed?

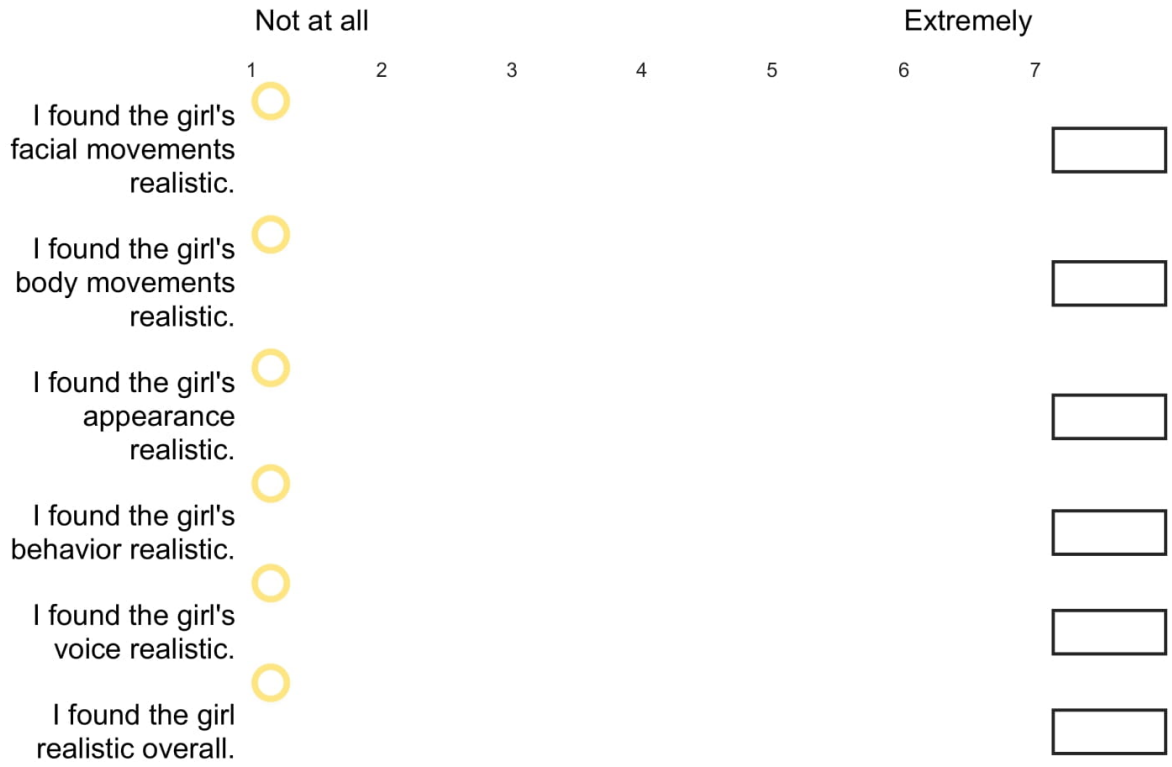
	Not at all						Extremely
	1	2	3	4	5	6	7
Intensity	◯						☐

How much does each statement apply for you?

	Not at all						Extremely
	1	2	3	4	5	6	7
It feels as if I am in the presence of another person in the room with me.	◯						☐

How much does each statement apply for you?

How much does each statement apply for you?



Photorealistic Appearance - Condition Angry - Full Anim

Please watch the following video:

What was the topic the virtual human was talking about?

- ◯ Forgetting important documents
- ◯ Being late
- ◯ Going out for lunch

Which emotion did you perceive in the virtual human?

- ◯ Angry
- ◯ Happy
- ◯ Sad
- ◯ Neutral
- ◯ Other

How intense was the emotion that you observed?

	Not at all						Extremely
	1	2	3	4	5	6	7
Intensity	○	○	○	○	○	○	○

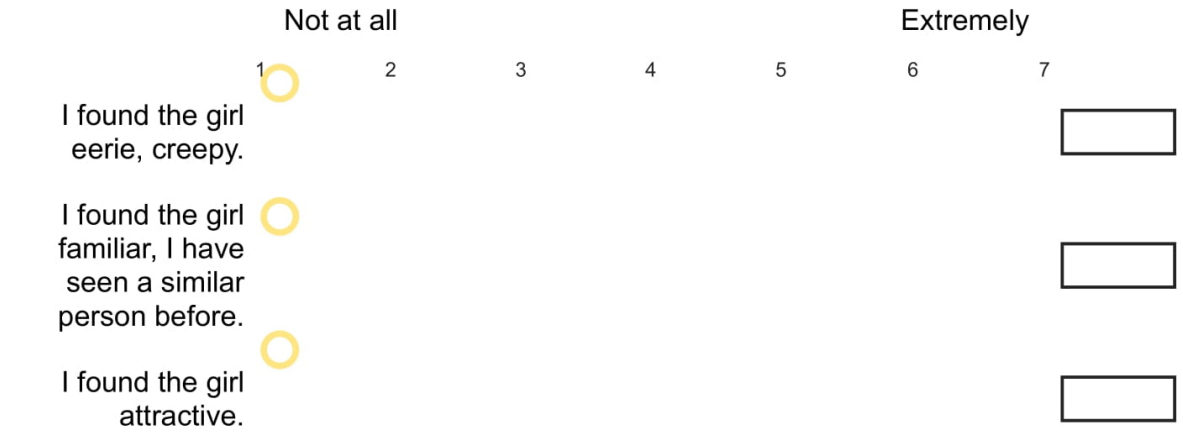
Not at all Extremely

How much does each statement apply for you?

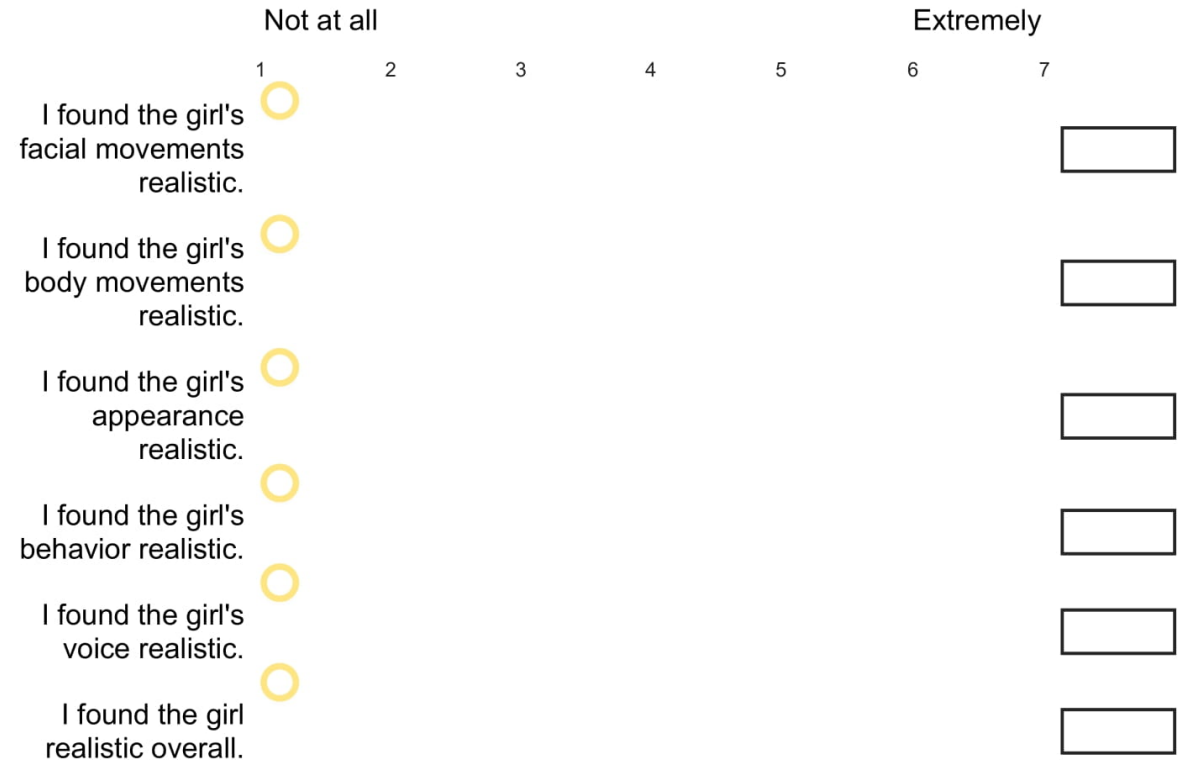
	Not at all						Extremely
	1	2	3	4	5	6	7
It feels as if I am in the presence of another person in the room with me.	○	○	○	○	○	○	○
It feels as if the girl is watching me and is aware of my presence.	○	○	○	○	○	○	○
The thought that the girl isn't real crossed my mind often.	○	○	○	○	○	○	○
The girl appears to be alive.	○	○	○	○	○	○	○
The girl is only a computerized image, not a real person.	○	○	○	○	○	○	○

How much does each statement apply for you?

	Not at all						Extremely
	1	2	3	4	5	6	7
I found the girl appealing, likable.	○	○	○	○	○	○	○



How much does each statement apply for you?



Photorealistic Appearance - Condition Sad - Full Anim

Please watch the following video:

What was the topic the virtual human was talking about?

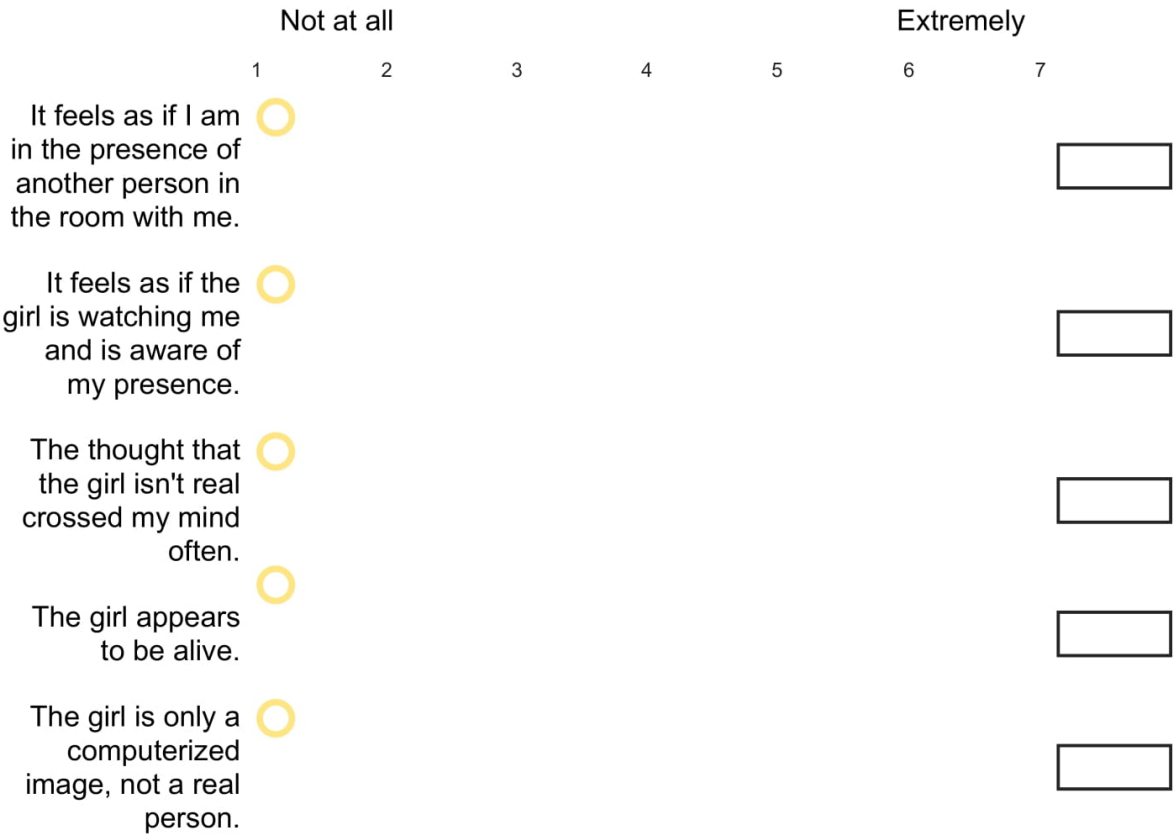
- ◯ Loneliness
- ◯ Forgotten birthday
- ◯ Car breaking down

Which emotion did you perceive in the virtual human?

- ◯ Angry
- ◯ Happy
- ◯ Sad
- ◯ Neutral
- ◯ Other

How intense was the emotion that you observed?

How much does each statement apply for you?



How much does each statement apply for you?